3. Точечные оценки и их свойства (несмещенность, эффективность, состоятельность).

Точечной статистической оценкой $\hat{\theta}$ параметра θ называется приближенное значение этого параметра, полученное на основе выборки. Оценка представляет собой число или точку на числовой прямой.

Оценка является функцией от результатов наблюдений:

$$\hat{\theta} = f(x_1, x_2, \dots, x_n)$$

Так как результаты наблюдений x_i являются случайными величинами, то и сама оценка $\hat{\theta}$ также является **случайной величиной**, обладающей своим математическим ожиданием, дисперсией и законом распределения.

Чтобы оценка считалась «хорошим» приближением к параметру, она должна обладать свойствами **состоятельности, несмещенности и эффективности**.

1. Состоятельность

Определение: Оценка $\hat{\theta}$ называется **состоятельной**, если при неограниченном увеличении объема выборки ($n \to \infty$) она стремится по вероятности к оцениваемому параметру θ:

$$\lim_{n\to\infty} P(|\hat{\theta} - \theta| < \epsilon) = 1$$

где $\epsilon > 0$ — произвольно малое число.

Смысл: Это свойство означает, что с ростом объема исходной информации вероятность значительного отклонения оценки от истины стремится к нулю. Если оценка не обладает состоятельностью, увеличение выборки не гарантирует уточнения значения параметра.

2. Несмещенность

Определение: Оценка $\hat{\theta}$ называется **несмещенной**, если ее математическое ожидание равно оцениваемому параметру θ при любом фиксированном объеме выборки n:

$$M(\hat{\theta}) = \theta$$

Если это равенство не выполняется, оценка называется **смещенной**, а разность $M(\hat{\theta}) - \theta$ называется **смещением**.

Смысл: Требование несмещенности гарантирует отсутствие систематических ошибок при оценивании. Среднее значение всех возможных оценок, вычисленных по разным выборкам одного объема, будет в точности равно истинному параметру.

Проиллюстрируем это свойство графически. Допустим, имеется два алгоритма расчета оценок характеристики θ:

$$A = f_1(x_1, x_2, \ldots, x_n) \quad \text{и} \quad B = f_2(x_1, x_2, \ldots, x_n).$$

Значения этих оценок изобразим точками на прямой (рис.3.1). Так как среднее значение оценки A совпадает с θ (полученные оценки расположены симметрично, относительно θ), то A — несмещенная оценка параметра θ. А оценка B — смещенная, т.к. ее среднее значение $M(B)$ не совпадает с θ.

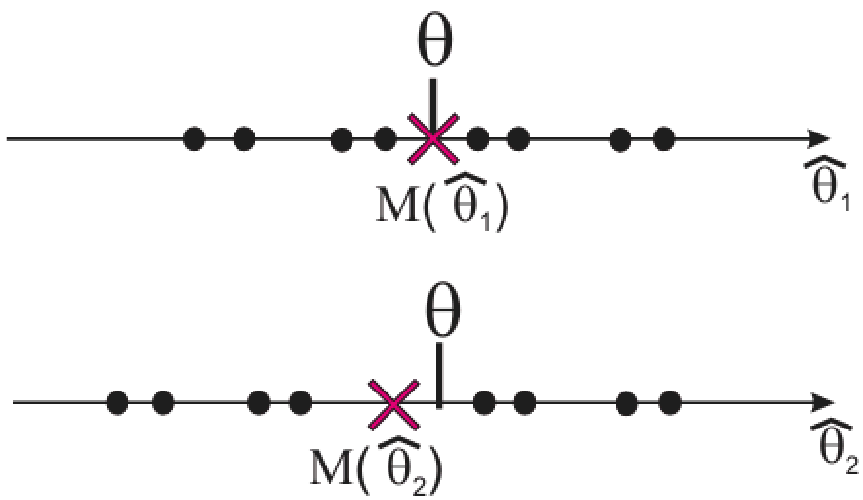

Рис.3.1.Иллюстрация свойства несмещенности оценок

3. Эффективность

Определение: Несмещенная оценка $\hat{\theta}$ называется **эффективной**, если она имеет минимальную дисперсию по сравнению с другими возможными несмещенными оценками того же параметра:

$$D(\hat{\theta}) = \min_{\hat{\theta}_i \in \Omega} D(\hat{\theta}_i)$$

где Ω — множество всех возможных несмещенных оценок параметра θ.

Предположим, что имеются две оценки параметра θ, рассчитанные на основе одной и той же информации и обе оценки - несмещенные. Естественно, считать лучшей ту оценку, у которой дисперсия (разброс относительно θ) меньше. Оценка A является более эффективной, чем оценка B (рис 3.2).

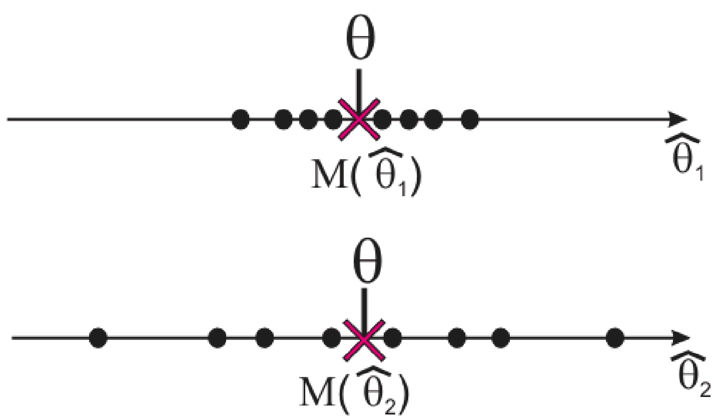

Рис.3.2.Иллюстрация свойства эффективности оценок

Для проверки эффективности используется **неравенство Рао — Крамера — Фреше**, которое задает нижнюю границу дисперсии:

$$D(\hat{\theta}) \geq \frac{1}{n \cdot I(\theta; x)}$$

где $I(\theta; x)$ — **информация Фишера** о параметре θ, содержащаяся в одном наблюдении. Для непрерывной величины она равна:

$$I(\theta; x) = \int_{-\infty}^{\infty} \left(\frac{\varphi'_\theta(x; \theta)}{\varphi(x; \theta)} \right)^2 \varphi(x; \theta) dx$$

3.2. Неравенство Рао-Крамера-Фреше

Пусть $\varphi(x,\theta)$ — плотность вероятностей случайной величины X, если X — непрерывная случайная величина, и $\varphi(x_i,\theta) = P(X = x_i, \theta)$ — вероятность того, что с.в. X приняла значение x_i, если X - дискретная случайная величина.

Если $\varphi(x,\theta)$ удовлетворяет условиям регулярности:

1. область возможных значений исследуемой случайной величины, в которой
   $\varphi(x,\theta) \neq 0$, не зависит от θ;
2. в выражении $M(\theta) = \int \theta \cdot L(x_1, \ldots, x_n, \theta) dx_1 dx_2 \ldots dx_n$ и тождестве $\int L(x_1, \ldots, x_n, \theta) dx_1 dx_2 \ldots dx_n = 1$ допустимо дифференцирование по θ под знаком интеграла;
3. величина $I(\theta, x) \neq 0$,

тогда для любой оценки $\widehat{\theta}$ неизвестного параметра θ имеет место неравенство Рао-Крамера-Фреше:

$$D(\widehat{\theta}) \geq \frac{1}{n \cdot I(\theta, x)}, \tag{3.5}$$

где $D(\widehat{\theta})$ - дисперсия оценки $\widehat{\theta}$ параметра θ; $I(\theta, x)$ - количество информации Фишера о параметре θ, содержащееся в единичном наблюдении:

$$I(\theta, x) = M\left[(\ln \varphi(x,\theta))'_\theta\right]^2. \tag{3.6}$$

По правилу дифференцирования сложной функции

$$(\ln y(x))'_x = \frac{y'(x)}{y(x)}.$$

Информация Фишера $(I(\theta; x))$ — это мера количества информации о неизвестном параметре θ, которую несёт в себе наблюдение x, отражающая чувствительность функции правдоподобия к изменениям параметра.

Она количественно оценивает, насколько хорошо данные позволяют "узнать" параметр модели, который их породил.

Что она характеризует?

1. **Плотность информации:** Чем больше значение $I(\theta; x)$, тем больше информации содержится в наблюдении x об θ, и тем более точно можно оценить этот параметр.
2. **Точность оценки (Неравенство Крамера-Рао):** Информация Фишера определяет нижний предел для дисперсии любого несмещенного оценщика параметра θ, т.е. $Var(\hat{\theta}) \geq 1/I(\theta; x)$.
3. **Чувствительность модели:** Она показывает, насколько быстро меняется функция правдоподобия при изменении θ. Если $I(\theta; x)$ велико, значит, небольшие изменения θ приводят к заметным изменениям в правдоподобии, что облегчает оценку.

Поэтому в случае дискретной случайной величины

$$I(\theta, x) = \sum_{i=1}^{n} \left[\frac{\varphi'(x_i, \theta)}{\varphi(x_i, \theta)} \right]^2 \cdot \varphi(x_i, \theta).$$

Если дисперсия оценки достигает этой нижней границы, она является эффективной.

Неравенство (5.11) позволяет найти тот минимум дисперсии, который должна иметь дисперсия оценки $D(\widehat{\theta})$, чтобы быть эффективной оценкой, т.е.

$$D_{\text{эф}}(\widehat{\theta}) = \min D(\widehat{\theta}) = \frac{1}{n \cdot I(\theta, x)}.$$

Аналогия:

Представьте стрельбу по мишени.

- **Несмещенность** — ваш прицел не сбит, и центр кучности попаданий совпадает с «десяткой».
- **Эффективность** — ваши выстрелы ложатся очень кучно, с минимальным разбросом.
- **Состоятельность** — чем больше вы тренируетесь (увеличиваете выборку), тем неизбежнее ваши попадания стягиваются в одну точку в самом центре.