

Таким образом, квадрат нормы погрешности разбивается в сумму квадрата нормы систематической погрешности

$$\delta_n^2 = \|p_n(\cdot) - p(\cdot)\|^2, \quad (12.11)$$

вызванной отклонением $p(\cdot)$ от $\mathcal{E}_n = \text{Span}(\varphi_1, \dots, \varphi_n)$, и квадрата нормы случайной ошибки уже в \mathcal{E}_n

$$\zeta_{nN}^2 = \|\pi_{nN}^*(\cdot) - p_n(\cdot)\|^2 = [\alpha_1^* - a_1]^2 + \dots + [\alpha_n^* - a_n]^2. \quad (12.12)$$

(или в пространстве параметров — коэффициентов разложения по базису $\varphi_1, \dots, \varphi_n$). Из разложения (12.10) ясно, что подобрав априори достаточно хорошо приближающее \mathcal{P} подпространство \mathcal{E}_n и сделав затем достаточно большое число N наблюдений, мы по формулам (12.7) и (12.9) построим проекционную оценку $\pi_{nN}^*(\cdot)$ неизвестной плотности со сколь угодно малой в метрике $\mathcal{L}^2(r)$ ошибкой. В принципе, в этих условиях существует такая зависимость $n = v(N)$, что при числе наблюдений $N \rightarrow \infty$ $\pi_{v(N), N}^*(\cdot)$ сходится к $p(\cdot)$ в норме $\mathcal{L}^2(r)$ почти наверное. Это следует из усиленного закона больших чисел Колмогорова.

Погрешность оценки неизвестной плотности проще всего охарактеризовать математическим ожиданием квадрата нормы погрешности. При таком подходе $\|\pi(\cdot) - p(\cdot)\|^2$ задает функцию потерь, а $E_p \|\pi^*(\cdot) - p(\cdot)\|^2$ становится функцией риска $\mathfrak{R}_{\Pi(N)}(P)$ для оценки $\pi^*(\cdot) = \pi(\cdot | \Xi_N)$. В регулярных ситуациях квантили $\Omega_\varepsilon(\|\pi^*(\cdot) - p(\cdot)\|)$ порядка $1 - \varepsilon$ с достаточно малым ε совпадают по порядку величины со средним квадратичным отклонением. В одну сторону это следует из неравенства Чебышева

$$\Omega_\varepsilon(\zeta^2) \leq \varepsilon^{-1} E \zeta^2;$$

в другую, когда $E \zeta^4 \leq \theta (E \zeta^2)^2$, то при $\varepsilon < \theta/4$

$$\Omega_\varepsilon(\zeta^2) \geq 2^{-1} E \zeta^2.$$

Возникает следующая задача. Пусть функции $\varphi_1(\cdot), \dots, \varphi_n(\cdot), \dots$ образуют ортонормированный базис пространства $\mathcal{L}^2(r)$. С каким n выгоднее всего по независимой выборке объема N построить проекционную оценку плотности $p(\cdot)$ ее распределения? В этой задаче для каждого n

$$E_p^{(N)} \|\pi_{nN}^* - p\|^2 = \sum_{j=n+1}^{\infty} |a_j|^2 + \frac{1}{N} \sum_{k=1}^n \text{Var}_p \alpha_k. \quad (12.13)$$

Для гладких плотностей и хороших базисов коэффициенты a_k убывают, хотя и не обязательно монотонно. Дисперсии же $\text{Var} \alpha_k$ примерно одинаковы, кроме, может быть, первой, если $\varphi_1(\omega) = 1(\omega)$, когда $\text{Var} \alpha_1 = 0$. Ясно, что (почти) оптимальное значение n должно (почти) минимизировать сумму (12.13). Если скорость убывания a_k известна, $|a_k| < C \cdot f(k)$, где $f(t) \rightarrow 0$