Ликбез по методу максимального правдоподобия!

Идея метода:
Имеется n наблюдений, значений случайных величин. Параметр θ не известен. Считаем вероятность получить имеющиеся наблюдения. Она зависит от θ. Выбираем $\hat{\theta}$ так, чтобы вероятность имеющихся наблюдений была наибольшей.

Пример с вероятностью.

Пример с функцией плотности.

Чем хорош метод максимального правдоподобия?

Теорема: **Оценки ML «хорошие»**:
Оценки будут:
Состоятельными:
$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \epsilon) = 0$
Асимптотически несмещенными:
$\lim E(\hat{\theta}_n) = \theta$
Асимптотически нормальными.
$\hat{\theta}_n \sim N$

Информацией Фишера (информационной матрицей) называют $I = -E(l''(\theta))$
Теоремка: $I = -E(l''(\theta)) = E(l'(\theta) \cdot l'(\theta))$

Теорема: **Неравенства Крамера-Рао** У любой другой несмещенной оценки дисперсия не меньше, чем I^{-1}

Теорема: **Оценки ML «лучшие»**:
Оценки ML являются асимптотически эффективными, т.е. их дисперсия является асимптотически минимально возможной и равна I^{-1}
Асимптотически эффективными среди асимпотически нормальных несмещенных.
Если взять другую асимпотически несмещенную нормальную оценку $\hat{\theta}_n^{alt}$, то у нее будет выше дисперсия, $Var(\hat{\theta}_n^{alt}) \geq Var(\hat{\theta}_n^{ML})$

Подведем итог: $\hat{\theta} \sim N(\theta, I^{-1})$

Настояющая I зависит от неизвестного θ
Однако можно оценить $\hat{I}$ по принципу: $\hat{I} = -l''(\hat{\theta})$
Значит $\widehat{Var}(\hat{\theta}) = \hat{I}^{-1}$

С помощью ML можно проверять гипотезы:
$H_0 : \theta = \theta_0$ Святая троица ML тестов
Тест Вальда (Wald test):

$W = (\hat{\theta} - \theta_0) \cdot I \cdot (\hat{\theta} - \theta_0)$

Score test (Lagrange multiplier test):
$LM = l'(\theta_0) \cdot I^{-1} \cdot l'(\theta_0)$

Likelihood ratio test:
$LR = -2(l(\theta_0) - l(\hat{\theta}))$

Если нулевая гипотеза верна, то все три статистики имеют χ^2 распределение с 1-ой степенью свободы.

Комментарий: формулы записаны в таком странном виде, чтобы были похожи на многомерный случай.

Прочие факты:

Факт 1. $E(l(\theta))$ достигает своего максимума при истинном θ

Факт 2. $Var(l'(\theta)) = I(\theta)$

Многомерный случай...
Имеет k параметров $\theta = (\theta_1, \ldots \theta_k)^t$
Проверяется гипотезы, состоящая из j ограничений.
В этом случае произойдут такие изменения:
Вместо производной $l'(\theta)$ будет вектор градиент $grad(l)$
Вместо второй производной $l''(\theta)$ будет матрица Гессе $H(\theta)$
Вместо I будет матрица
Оценка $\hat{I}$ считается как $\hat{I} = -H(\hat{\theta})$
Теоремка примет вид: $I = -E(H(\theta)) = E(grad(l)^t grad(l))$
Изменятся формулы трех статистик.

Статистики будут иметь χ^2_j распределение

Примечания мелким шрифтом:
Указанные теоремы верны при соблюдении следующих технических условий:

Более сильный вариант неравенства Крамера-Рао имеет вид:

Доказательства можно найти, например, в