Плотность распределения измеряемой величины, основанная на имеющейся информации

В.Воегер, ПТБ, Германия

Wolfgang.Woeger@ptb.de

Современный подход к оценке измерительной информации, рекомендованный Руководством по выражению неопределенности измерений (Guide to the Expression of Uncertainty in Measurement (GUM), основан на математической формулировке простой идеи, что любого рода имеющаяся информация, используемая для получения оценки измеряемой величины, генерирует соответствующее представление об измеряемой величине. После измерения знание измеряемой величины никогда не бывает полным, поскольку предоставленная информация не позволяет установить значение измеряемой величины точно. Правильнее сказать, всегда присутствует неопределенность в знании значения измеряемой величины. Математически знание об измеряемой величине представляется в виде распределения вероятностей на множестве возможных значений измеряемой величины. В данном случае вероятность того, что значение измеряемой величины совпадет с выбранным возможным значением интерпретируется как степень ожидания того, что это событие реализуется. Степень ожидания (вероятность) соответствует всей научно обоснованной информации об измеряемой величине. В метрологии обычно данная информация состоит из статистических данных (результатов измерений) и нестатистических данных, таких как, например, границы систематических отклонений. Интерпретация вероятности, принятая в данном подходе, не может быть унифицирована с общепринятой частотной интерпретацией вероятности. Она более общая, гибкая и соответствует высказыванию Эйнштейна, что наука это не более чем здравый смысл аргументированный соответствующим образом. Применяя принцип максимума энтропии и теорему Байеса, мы можем найти плотность распределения вероятностей (pdf) для измеряемой величины, соответствующую всей представленной, но неполной информации.

Следует понимать, что pdf представляет оценку измеряемой величины на основе имеющейся информации об измеряемой величине и, следовательно, никогда не может интерпретироваться как частотное распределение наблюдаемых значений. Более того, помимо информации о систематических эффектах статистический разброс результатов является дополнительной информацией, которая будет учтена в pdf для измеряемой величины. В отличие от плотности распределения результатов измерений, вид и параметры pdf для измеряемой величины известны. Если через ξ обозначить возможные значения измеряемой величины, а через X неизвестное значение измеряемой величины, то математическое ожидание

$$x = \mathrm{E}X = \int \xi f_X(\xi | I)\mathrm{d}\xi$$

распределения вероятностей $f_X(\xi | I)$ (pdf) является наилучшей оценкой измеряемой величины (результатом измерения).

Обозначение для плотности распределения явно учитывает тот факт, что каждая и все плотности распределения обусловлены имеющейся информацией о величине X. Дисперсия

$$u^2(x) = \mathrm{Var}(X) = \int (\xi - x)^2 f_X(\xi | I)\mathrm{d}\xi$$

распределения вероятностей это квадрат стандартной неопределенности, связанной с результатом измерения. Таким образом, стандартная неопределенность является параметром pdf. Она относится к неопределенности в знании измеряемой величины и не выражает вероятностного утверждения относительно измеряемой величины, как в случае так называемой расширенной неопределенности. Интегралы в двух вышеприведенных выражениях могут быть достаточно сложными. В таких случаях следует использовать компьютеры. Соответствующие численные методы будут представлены М.Коксом и Б.Зибертом на данном семинаре.

Хотя это явно и не сформулировано в GUM, концепция pdf лежит в основе всех рекомендаций этого документа. Эта концепция в основном своим появлением обязана успешным исследованиям Бернулли, Байеса и Лапласа. Описанный подход к обработке информации заменяет традиционный анализ погрешностей, основанный на частотной интерпретации вероятности, который долгое время применялся в метрологии. В силу объективных причин анализ погрешностей не позволяет обращаться с систематическими отклонениями также как со случайными, в прошлом это неизбежно вело к противоречиям и заключениям о неопределенностях, которые часто на практике имели неправдоподобные значения. Описанный Байесовский подход позволяет избежать все недостатки анализа погрешностей. В зачастую неправдоподобном случае, когда единственный источник неопределенности представлен только статистической информацией, эти два подхода дают на практике численно одинаковые результаты.

В данном выступлении будет рассмотрен вопрос сопоставления pdf неизвестной величине на основе информации более или менее общего вида (принцип максимума энтропии). Также будет кратко рассмотрен случай статистической информации в виде серии повторных измерений в условиях повторяемости (теорема Байеса). Будут представлены примеры.

Предполагается, что в общем случае информация о величине X состоит из известных оценок γ_j функций $G_j(X)$. Эти оценки должны входить в плотность распределения таким образом, чтобы $EG_j(X)=\gamma_j$. Среди многих плотностей вероятностей, соответствующих данной информации, вполне естественно выбрать ту единственную, которая включает минимум информации. Мерой потери информации, содержащейся в плотности вероятности, является энтропия, предложенная Шенноном (1948)

$$S=-\int_{\Omega} f_X(\xi)\ln f_X(\xi)\,\mathrm{d}\xi$$

Соответственно, принцип максимума энтропии – сформулированный Джайенсом, 1957 - состоит в максимизации этого функционала при условии следующих ограничений $EG_j(X)=\gamma_j$.

В случае если информация о виде частотного распределения вероятностей представлена рядом выборочных значений, pdfs для неизвестных параметров частотного распределения могут быть получены, используя теорему Байеса. Последняя является следствием основного правила перемножения условных вероятностей в теории вероятностей. Имея исходную оценку (априорную pdf) параметра, теорема объясняет, каким образом новая оценка (апостериорная pdf) получается на основе ряда наблюденных значений, извлеченных из частотного распределения. Формальное правило этого познавательного процесса может быть кратко записано как *апостериорное = правдоподобие × априорное.* Априорная плотность распределения может быть получена, применяя принцип максимума энтропии к общей информации, которая представлена без статистических данных. Типичным примером является проблема повторной калибровки.