Conceptually, law-like rules came from the philosophy of science. These rules attempt to mathematically capture the essential features of scientific laws:

(1) high level of generalization;
(2) simplicity (Occam s razor); and,
(3) refutability.

Formally, an IF-THEN rule C is presented as $A_1\&\quad\&A_k \Rightarrow A_0$, where the IF-part, $A_1\&\ldots\&A_k$, consists of true/false logical statements $A_1, \quad ,A_k$, and the THEN-part consists of a single logical statement A_0. Rule C allows one to generate sub-rules with a truncated IF-part, e.g. $A_1\&A_2 \Rightarrow A_0$, $A_1\&A_2\&A_3 \Rightarrow A_0$ and so on. It is known that a sub-rule is logically stronger than the rule used to construct the sub-rule. Thus, if some rule and its sub-rule C classify correctly the same set of examples, then the sub-rule is preferred.

In general, there are three reasons to prefer the sub-rule:

1. The sub-rule is more general (logically stronger and describes the same set of events).
2. The sub-rule is simpler then the rule, because it consists of fewer statements in the IF-part.
3. Sub-rule is better testable (more refutable) then the rule, because the larger set of possible examples may falsify it (the IF-part of the sub-rule is less restrictive).

Thus, if a rule covers the set of examples then one can test that no one of its sub-rules also covers the same set of examples. Otherwise, this sub-rule or maybe some of its sub-rules will be preferred, because this sub-rule is simpler, more general and more refutable.

In deterministic case, a *law-like* rule can be defined (for some set of examples) as a rule without sub-rules covering this set of examples. In other words *law-like* rule is the rule which is true for some set of examples, but no one of its sub-rule is true for this data.

If examples contain noise, which is typical in many fields, the probabilistic characteristics of the expressions are used instead of crisp (true/false) values. The conditional probability of the rule is used in the MMDR method as this charateristic. For rule C, its conditional probability Prob(C) = Prob(A0/A1&...&Ak) is defined, assuming that Prob(A1&...&Ak) > 0. Similarly conditional probabilities Prob(A0/ Ai1&...&Aih) are defined for sub-rules Ci, such as $Ai1\&\quad\&Aih \Rightarrow A0$., assuming that Prob(Ai1&...&Aih) > 0. Conditional probability, Prob(C) = Prob(A0/A1&...&Ak), is used for estimating forecasting power of the rule to predict A0. In addition, the conditional probability is a major tool for defining non-deterministic (probabilistic) law-like rules (regularities) (Vityaev E et al. 1998;1992).

The rule is a probabilistic *law-like* rule iff all of its sub-rules have a statistically significant lower conditional probability than the rule. Another definition of *law-like* rules can be given in terms of generalization. The rule is *law-like* iff it can not be generalized without producing a statistically significant reduction in its conditional probability. Law-like rules defined in this way hold all three listed above properties (properties of scientific laws), i.e., these rules are:

(1) general from a logical perspective,
(2) simple, and
(3) refutable.