SUMMARY OF BASIC RULES AND NOTATION
(Continued from inside front cover)

Continuous Distributions: If x is continuously variable, we denote the probability, given A, that it lies in the range (x, x+dx) by

$$(dx|A) = (x|A)\ dx$$

Thus the same bracket symbols (|) are used for probabilities and probability densities. This causes no confusion, since the distinction is determined by whether the quantity is discrete or continuous. Rule 1 and Bayes' theorem then take the same form as above, since dA and/or dB cancel out; and the summations above become integrations.

Prior Probabilities: The initial information available to the robot at the beginning of any problem is denoted by X. $(A|X)$ is then the prior probability of A. Applying Bayes' theorem to take account of new evidence E yields the posterior probability $(A|EX)$. In a posterior probability we sometimes leave off the X for brevity: $(A|E) \equiv (A|EX)$.

Prior probabilities are determined by Rule 4 when applicable; or more generally by the principle of maximum entropy (Lect. 10): choose the $p_i \equiv (A_i|X)$ so as to maximize $H = - \Sigma_i\ p_i \log p_i$ subject to constraints represented by X. In the continuous case this becomes: maximize $H = - \int p(x) \log [p(x)/m(x)]\ dx$, where the measure m(x) is determined by invariance under the group of transformations which convert the problem into an equivalent one, for consistency in sense (b) above (Lect. 12).

Decision Theory: (Lect. 13). Enumerate the possible decisions $D_1 \ldots D_k$ and introduce a function $L(D_i, \theta_j)$ representing the "loss" incurred by making decision D_i if θ_j is the true state of nature. Make that decision D_i which minimizes the expected loss $\langle L \rangle_i = \Sigma_j\ L(D_i, \theta_j)(\theta_j|EX)$ over the posterior distribution of θ_j.

Probability and Frequency: The above rules are shown to apply to general inductive inferences, whether or not any random experiment is involved. Many applications can be carried to completion without ever mentioning frequencies (Lectures 5,6,8,9,11,14,18).

If a problem does involve a random experiment, connections between probability and frequency will appear as mathematical consequences of the theory. Most random experiments are exchangeable sequences (Lect. 17); here the probability of an event is numerically equal to the estimate of frequency which minimizes the expected square of the error. Conversely, if an experiment has been repeated many times, the probability of any event at the next trial approaches its observed frequency (Lect. 16).

Probabilitees derived from maximum entropy subject to constraints are equal to the frequencies which can be realized in the greatest number of ways subject to the same constraints (Lect. 10). Probabilities derived by invariance under a transformation group are equal to the frequencies most likely to be produced in the sense that they require the least "skill" (Lect. 12).