Probability Measures

Statistics 110

Summer 2006

Probability Measures

A **probability measure** on a sample space Ω is a function P from subsets of Ω to the real numbers satisfying the following axioms:

1. $P[\Omega] = 1$ (Boundedness)

2. If $A \subset \Omega$, then $P[A] \geq 0$ (Positivity)

3. If A_1 and A_2 are disjoint (i.e. $A_1 \cap A_2 = \phi$), then

$$P[A_1 \cup A_2] = P[A_1] + P[A_2]$$

More generally, if $A_1, A_2, \ldots, A_n, \ldots$ are disjoint then

$$P\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} P[A_i]$$

This is known as countable additivity. Finite additivity (i.e. $n < \infty$) is a special case of this.

The first axiom states that the chance of seeing one of the possible outcomes is 1. This matches with the equally likely cases discussed earlier where

$$P[\Omega] = \frac{\#\text{ outcomes in }\Omega}{\#\text{ outcomes in }\Omega} = 1$$

The second axiom states that probabilities are non-negative. Again this matches with the equally likely case from earlier since

$$0 \leq \#\text{ outcomes in } A \leq \#\text{ outcomes in } \Omega$$

Kolmogorov (1933) found that all probability calculations and theorems can be obtained by the systematic application of these three axioms.

These serve as axioms for the mathematical development of all probability theory, including:

- Continuous random variables
- Limit theorems
- Stochastic processes

From these three axioms, we can get a number of other useful properties

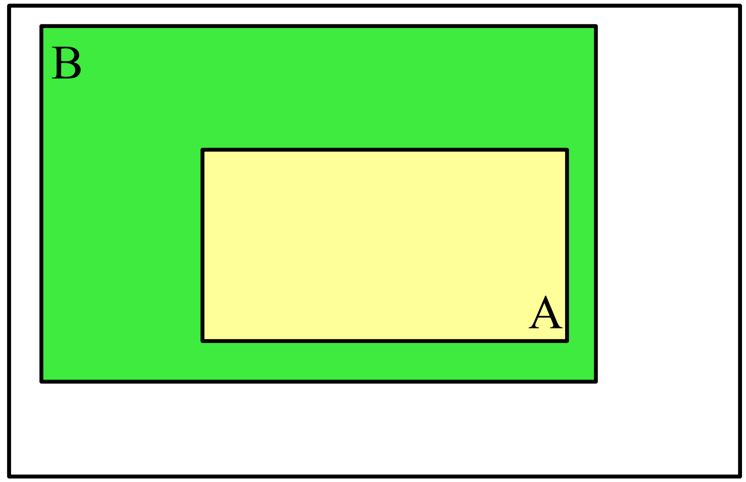
1. $P[A^c] = 1 - P[A]$ (Complement Rule)

$$P[A] + P[A^c] = P[\Omega] = 1$$

2. $P[\phi] = 0$

$$P[\phi] = P[\Omega^c] = 1 - P[\Omega] = 0$$

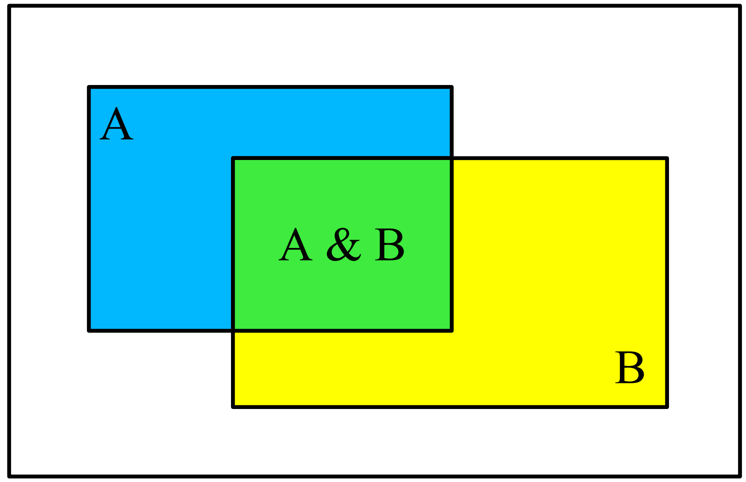
3. If $A \subset B$ then $P[A] \leq P[B]$



$$B = (A \cap B) \cup (A^c \cap B)$$

$$P[B] = P[A \cap B] + P[A^c \cap B] = P[A] + P[A^c \cap B] \geq P[A]$$

4. $P[A \cup B] = P[A] + P[B] - P[A \cap B]$ (Addition Rule)



$$A \cup B = (A \cap B^c) \cup (A^c \cap B) \cup (A \cap B)$$

$$P[A] = P[A \cap B] + P[A \cap B^c]$$
$$P[B] = P[A \cap B] + P[A^c \cap B]$$

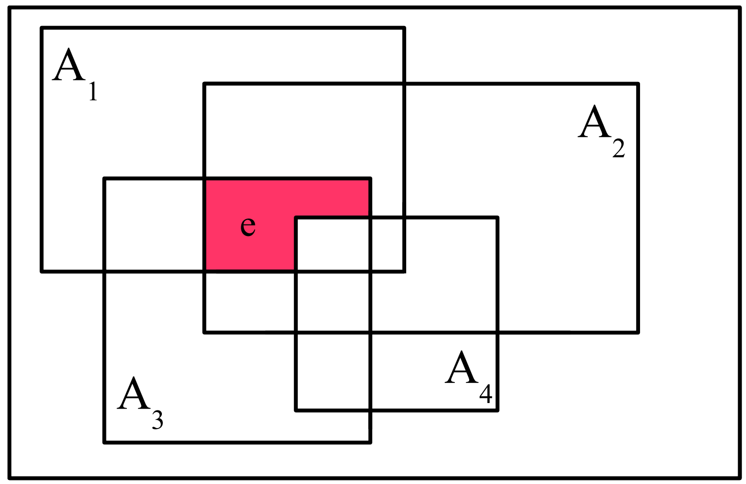
The addition rule can be extended to an arbitrary number of events as follows. For a collection of events $A_1, A_2, \ldots, A_n$, (n possibly ∞) let

$$
\begin{array}{ll}
p_i = P[A_i] & S_1 = \sum_i p_i \\
p_{ij} = P[A_i \cap A_j] & S_2 = \sum_{i<j} p_{ij} \\
p_{ijk} = P[A_i \cap A_j \cap A_k] & S_3 = \sum_{i<j<k} p_{ijk}
\end{array}
$$

Then

$$
P\left[\bigcup_{i=1}^{n} A_i\right] = S_1 - S_2 + S_3 - S_4 \ldots \pm S_n
$$

Proof. Consider and outcome $e \in \bigcup_{i=1}^{n} A_n$ and suppose that e is in $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ but not any other A_i.



(In this case e is in A_1, A_2, and A_3.) Let $\alpha = P[\{e\}]$ and consider the contribution of α to the RHS.

- In S_1, α is counted $\binom{k}{1}$ times.
- In S_2, α is counted $\binom{k}{2}$ times.
- In S_k, α is counted $\binom{k}{k}$ times.

Hence α is counted in the RHS

$$m = \binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \ldots \pm \binom{k}{k}$$

times. From the binomial expansion

$$0 = (1-1)^k = 1 - \left\{ \binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \ldots \pm \binom{k}{k} \right\}$$

so m is 1. $\square$

Theorem. *Binomial Theorem*

$$(x+y)^n = x^n + \binom{n}{1} x^{n-1} y + \binom{n}{2} x^{n-2} y^2 + \ldots + y^n$$

Example: The matching problem

Two equivalent decks of N distinct cards are each randomly ordered. Then they are matched against each other. What is the probability there are no matches (call it p_0).

- If N is 3,

$$\begin{array}{cccc} \text{Deck 1} & & \text{Deck 2} & \\ 1\;2\;3 & \mathbf{1}\;\mathbf{2}\;\mathbf{3} & \mathbf{1}\;3\;2 & 2\;1\;\mathbf{3} \\ & \boxed{2\;3\;1} & \boxed{3\;1\;2} & 3\;\mathbf{2}\;1 \end{array}$$

  so $p_0 = \frac{1}{3}$.

- What if N is 30,000?

$$\begin{array}{cccccc} \text{Deck 1}: & 1 & 2 & 3 & \ldots & N \\ \text{Deck 2}: & x_1 & x_2 & x_3 & \ldots & x_N \end{array}$$

We can assume without loss of generality (WLOG) that Deck 1 is always in numerical order. (If it isn't, do the same switches to both decks so that the first ends up in numerical order.) Also assume that each possible outcome for deck 2 has the same probability ($=\frac{1}{N!}$).

$$P[\text{No Match}] = 1 - P[\text{At least one match}]$$

Let A_i be the event {match at position i} $= \{x_i = i\}$.

Then

1.

$$\begin{aligned} P[\text{At least one match}] &= P\left[\bigcup_{i=1}^{N} A_i\right] \\ &= S_1 - S_2 + S_3 - S_4 \ldots \pm S_N \end{aligned}$$

2. $P[A_iA_j] = P[A_1A_2] = \frac{(N-2)!}{N!}, P[A_iA_jA_k] = P[A_1A_2A_3] = \frac{(N-3)!}{N!}$, etc.

So

$$S_1 = \sum_i P[A_i] = NP[A_1] = N\frac{1}{N} = 1$$

$$S_2 = \sum_{i<j} P[A_iA_j] = \binom{N}{2} P[A_1A_2] = \binom{N}{2}\frac{(N-2)!}{N!} = \frac{1}{2!}$$

$$S_3 = \sum_{i<j<k} P[A_iA_jA_k] = \binom{N}{3} P[A_1A_2A_3] = \binom{N}{3}\frac{(N-3)!}{N!} = \frac{1}{3!}$$

Hence

$$P[\text{At least one match}] = 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots \pm \frac{1}{N!}$$

$$P[\text{No match}] = \frac{1}{2!} - \frac{1}{3!} + \ldots \pm \frac{1}{N!}$$

Recall that

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

$$e^{-1} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \dots$$

so that $P[\text{No match}] \approx e^{-1} \approx 0.367$.

Inequalities

Sometimes, exact probability statements aren't needed. Instead bounds are good enough. The following are a couple of useful ones

- Boole's inequality

$$P[A_1 \cup A_2 \cup \ldots] \leq P[A_1] + P[A_2] + \ldots$$

- Bonferroni's inequality

$$P[A_1 \cap A_2 \cap \ldots \cap A_n] \geq P[A_1] + P[A_2] + \ldots + P[A_n] - (n-1)$$

Boole's inequality is useful in making statements about rare events. If events $\{A_1, A_2, \ldots, A_n\}$ are unlikely individually, then the chance that any of them occur is still unlikely.

For example, suppose that each of us bought 1 lottery ticket for the next draw in Mass Millions (Pick 6 from 1 to 49). $P[\text{I win}] = \frac{1}{13{,}983{,}816}$ (or anybody else). If there are 14 people here,

$$P[\text{At least one of us wins}] \leq \frac{14}{13,983,816} \quad (\approx \frac{1}{1,000,000})$$

Bonferroni's inequality is often useful in making statement about common events. If events $\{A_1, A_2, \ldots, A_n\}$ are all likely individually, then the chance that **all** of them occur is still likely, though not as likely as each individually (as $\bigcap A_i \subset A_j$ for all j).

For example, a common statistical inference technique is the use of Confidence Intervals (CI). The procedure for generating CIs has the property that the $P[\text{True parameter value in interval}] = C$. Often in an analysis there are multiple intervals being generated. Suppose that there are n of them and let

$$A_i = \text{ Interval } i \text{ contains the true parameter value}$$

What is the probability that all intervals contain the true parameter values assuming that $n = 10$, and $C = 0.95$ (a popular choice).

$$P\left[\bigcap_{i=1}^{n} A_i\right] \geq 10 \times 0.95 - (10 - 1) = 0.5$$

So there could be a 50:50 chance of making at least one incorrect statement.

How big does C need to be for so that the probability that all intervals contain the true value is at least 0.95?

Need $10C - 9 \geq 0.95$ which implies $C \geq \frac{9.95}{10} = 0.995$

Also Bonferroni's inequality is sometimes written in the form

$$P\left[\bigcap_{i=1}^{n} A_i\right] \geq 1 - \sum_{i=1}^{n} P[A_i^c]$$

(to prove just replace $P[A_i]$ with $1 - P[A_i^c]$ in the original formulation)

How to define a probability measure?

The three axioms gives us a way of defining probability measures for countable (& finite) sample spaces (excludes cases such as $\Omega = [0, 1]$). Assume that $B_1, B_2, \dots$ are the elementary outcomes of the experiment. Then any set of numbers $p_i = P[B_i]$ satisfying

1. $0 \leq p_i \leq 1$

2. $\sum_{i=1}^{\infty} p_i = 1$

gives a valid probability measure.

Example: Tossing a biased coin. Assume that the probability of a head on each flip is $\frac{2}{3}$. Flip the coin until a tail appears and let X be the flip number when this occurs ($\Omega = \{1, 2, 3, \ldots\}$). Then $P[X = i]$ satisfies (assuming independence)

$$P[X = i] = \frac{1}{3}\left(\frac{2}{3}\right)^{i-1}; i = 1, 2, \ldots$$

Since $P[X = i]$ satisfies the 2 conditions (you should check this) it defines a valid probability measure.

However any other set of $\{p_i\}$ satisfying the 2 conditions could also be used as a probability model describing flipping this biased coin, though it probably would be a bad description (as it probably will violate the probability of getting a head on each flip of $\frac{2}{3}$.