Gov 2001 Section 1, Feb. 7, 2025

Naive Definition of Probability

Notice the condition of counting possibilities and applying the naive definition: *All outcomes are equally likely.* The textbook calls it "symmetry" (p.7). For example, consider when you choose to drink coffee (c), tea (t), or milk (m) once in the morning and once in the afternoon. We cannot count (c, t) and (t, c) as just one possible outcome, because some possible outcomes like (c, c) are counted only once and thus $P((c, t) \cup (t, c)) = 2P(c, c) = 2P(t, t)$. When the question says something like "a *fair coin*", "a *well-shuffled* deck of cards", or "a simple *random* sample", this is where you would find symmetry.

For multiplication rule, also notice its assumption: the number of outcomes in one experiment does not depend on the outcome of the other experiment. For questions like assigning a objects into b containers, you can start by listing out the ways of choosing from the objects/containers, and then calculate for each way, the number of ways to assign the containers/objects.

Permutations and combinations: A permutation is an arrangement of objects in some order, while a combination is where order does not matter. For example, there are $n \geq 3$ people running a race. To choose the 1st, 2nd, and 3rd place (permutation), there are $n(n-1)(n-2)$ ways. To choose the top 3 (combination), there are $\binom{n}{3}$ ways.

Below are some practice problems of symmetry and multiplication rule:

Q1. A manager has n employees and a list of n tasks. Instead of making sure that each employee receives exactly one task, the manager assigns the tasks completely randomly (so each task is equally likely to be assigned to any employee, independent of the other tasks).

(a) Find the probability that each employee is assigned exactly one task.

By multiplication rule, the denominator is n^n. To make sure each employee is assigned exactly one task, after assigning the first task to one employee, the second task can only be assigned to one of the $n-1$ employees. Following this logic, the probability is $\frac{n!}{n^n}$.

(b) Find the probability that at least one employee gets assigned no tasks.

This is just the complement of (a). So the probability is $1 - \frac{n!}{n^n}$.

(c) Find the probability that there is exactly one employee who gets assigned no tasks.

From (a), this event implies that the task of one random employee is assigned to another random employee. We have $\binom{n}{2}$ ways to choose the two "mismatched" employees. For each way, we can choose $\binom{n}{2}$ tasks to assign to this pair, and the two chosen tasks have two ways to be assigned (to either of the chosen employees). Then, we can assign the rest $n-2$ tasks to the rest $n-2$ employees, and from (a) we know there should be $(n-2)!$ ways. Therefore, the probability is

$$\frac{\binom{n}{2} \times \binom{n}{2} \times 2 \times (n-2)!}{n^n} = \frac{n(n-1)n!}{2n^n}$$

Another solution: we first choose the one who gets no tasks (n ways), and then choose the one who gets two tasks ($n-1$ ways). For each of the $n-1$ ways, we have $\binom{n}{2}$ ways of choosing the tasks to be assigned to this unfortunate person. And then assign the rest $n-2$ to $n-2$ as usual.

$$\frac{n \times (n-1) \times \binom{n}{2} \times (n-2)!}{n^n} = \frac{n(n-1)n!}{2n^n}$$

(d) Find the probability that there is one employee who gets assigned all the tasks.

There are n ways to choose which employee gets all the tasks, so the probability is

$$\frac{n}{n^n} = \frac{1}{n^{n-1}}$$

(e) Find the probability that there are two employees who, collectively, get assigned all the tasks (but neither of them individually gets assigned all the tasks).

There are $\binom{n}{2}$ ways to choose the two employees. For each way, there are 2^n ways to assign the tasks, while two of them assign all the tasks to one employee and should be excluded. Therefore, the probability is

$$\frac{\binom{n}{2} \times (2^n - 2)}{n^n} = \frac{(n-1)(2^n-2)}{2n^{n-1}}$$

Q2. Two gamblers are playing a coin toss game. Gambler A has $(n+1)$ fair coins. Gambler B has n fair coins. What is the probability that A will have more heads than B if they all flip their coins? (Hint: first consider what will happen if they both have n coins)

After A and B both flip their first n coins, the results could be either A has more heads (E_1), B has more heads (E_2), or they have the same number of heads (E_3). Let $P(E_1) = P(E_2) = x$ by symmetry, so $P(E_3) = 1 - 2x$ by the axiom of probability. The event A has more heads than B after flipping all happens when (1) E_1 happens; or (2) E_3 happens and A lands a head in the $n+1$ flip. Each flip is independent, so the probability is $x + (1-2x) \times 0.5 = 0.5$.

Conditional Probability

For Bayes rules and LOTP with extra conditioning, it is simple: just make sure the variable being conditioned on is also conditioned on for each term of the other side of the equation.

$$P(A \mid B, C) = \frac{P(B \mid A, C)\, P(A \mid C)}{P(B \mid C)}.$$

Let $\{B_i\}_{i=1}^n$ be a *partition* of the sample space (i.e., disjoint events whose union is the entire space). For an event A and a conditioning event C, the law of total probability states:

$$P(A \mid C) = \sum_{i=1}^{n} P(A \mid B_i, C)\, P(B_i \mid C).$$

Q3. (Monty Hall problem) You are a contestant on a game show and asked to choose one of three closed doors. Behind exactly one door is a car; behind the other two are goats. After you choose a door (but before it is opened), the host, who knows what is behind all the doors, opens one of the remaining doors to reveal a goat. You are then given the choice to stick with your original choice or to switch to the other unopened door. *Should you switch? Calculate the probability of winning the car for each strategy (switch or stay).*

This is a classic problem using LOTP. Let

$$A_i = \{\text{car is behind Door } i\}, \quad i = 1, 2, 3,$$
$$S = \{\text{event that you switch to the other unopened door}\}.$$

Since the car is equally likely to be behind any of the three doors, we have $P(A_1) = P(A_2) = P(A_3) = 1/3$. Now, we calculate the probability of winning if we switch, by LOTP:

$$P(\text{win} \mid S) \;=\; P(\text{win} \mid S, A_1)\, P(A_1) \;+\; P(\text{win} \mid S, A_2)\, P(A_2) \;+\; P(\text{win} \mid S, A_3)\, P(A_3).$$

- If the car is behind Door 1 $\big(A_1\big)$, then your original choice (Door 1) was correct. By switching, you *must* choose a goat. Hence:

$$P(\text{win} \mid S, A_1) \;=\; 0.$$

- If the car is behind Door 2 $\big(A_2\big)$, the host is forced to open Door 3 (to reveal a goat), and switching leads you to Door 2 (the car). Hence:

$$P(\text{win} \mid S, A_2) \;=\; 1.$$

- If the car is behind Door 3 $\big(A_3\big)$, by similar reasoning, the host opens Door 2, and switching leads you to Door 3 (the car). Hence:

$$P(\text{win} \mid S, A_3) \;=\; 1.$$

Putting these together:

$$P(\text{win} \mid S) = 0 \times \tfrac{1}{3} \;+\; 1 \times \tfrac{1}{3} \;+\; 1 \times \tfrac{1}{3} = \tfrac{2}{3}.$$

Therefore, **switching** gives a $\frac{2}{3}$ probability of winning the car.

Now, if you stick with your original choice, you rely entirely on having picked the correct door at the start. We have

$$P(\text{win} \mid S^c) = P(\text{car behind your chosen door}) = \tfrac{1}{3}$$

Therefore, it is strictly better to switch.

Q4. Consider a chip-producing machine that have a hard-to-detect defect. Chips are either good or defective. If the machine is defective, then the produced chips will be defective with probability c, independently. If the machine is not defective then the chips will all be good. Let p be the probability that the machine is defective. The machine is used to produce $n \geq 2$ chips.

(a) Find the probability that none of the n chips are defective.

Let the event that the i-th chip is good be C_i. Let $A_n = C_1 \cap C_2 \cap \cdots \cap C_n$. Let the event that the machine is good be M. We have, by LOTP,

$$P(A_n) = P(A_n|M)P(M) + P(A_n|M^c)P(M^c) = (1-p) + p(1-c)^n$$

(b) Find the probability that the second chip is good, given that the first chip is good.

From (a), we know that

$$P(C_2|C_1) = \frac{P(C_2 \cap C_1)}{P(C_1)} = \frac{(1-p)+p(1-c)^2}{(1-p)+p(1-c)}$$

(c) Find the probability that the machine is defective, given that the n chips are all good.

By Bayes' rule, we have

$$P(M^c|A_n) = \frac{P(A_n|M^c)P(M^c)}{P(A_n)} = \frac{p(1-c)^n}{(1-p)+p(1-c)^n}$$

(d) A random sample of k out of the n chips are tested (with all sets of k chips equally likely to be chosen). A tested chip either passes (i.e., the test says that the chip is good) or fails (i.e., the test says that the chip is defective). The test is imperfect though: a defective chip passes the test with probability t, with $0 < t < 1$. A good chip always passes the test. Find the probability that the machine is defective, given that all k tested chips pass.

Let the event that all k chip passes the test be $B_k = T_1 \cap T_2 \cap \cdots \cap T_k$. First, we calculate

$$\begin{aligned}
P(B_k|M^c) &= \prod_{j=1}^{k} P(T_j|M^c) \\
&= \prod_{j=1}^{k} [P(T_j|C_j, M^c)P(C_j|M^c) + P(T_j|C_j^c, M^c)P(C_j^c|M^c)] \\
&= (1-c+ct)^k
\end{aligned}$$

By Bayes' rule, we have

$$\begin{aligned}
P(M^c|B_k) &= \frac{P(B_k|M^c)P(M^c)}{P(B_k|M^c)P(M^c) + P(B_k|M)P(M)} \\
&= \frac{p(1-c+ct)^k}{p(1-c+ct)^k + 1 - p}
\end{aligned}$$

PMF, CDF, Bernoulli, Binomial

The support of a discrete random variable X is the set of values x such that $P(X = x) > 0$. For example, the support of $X \sim \text{Bern}(p)$ is $0, 1$ and the support of $Y \sim \text{Bin}(n, p)$ is $0, 1, \ldots, n$. To confirm that you got the PMF right, plug in some extreme values in the support (usually end points like 0, 1, or n to simplify calculation) into the function and see if the probability turns out correctly.

For common distributions, it is important to understand their real-world interpretations. A Bernoulli random variable $X \sim \text{Bern}(p)$ models a single trial that can result in a success or a failure, with the probability of success p. A Binomial random variable $Y \sim \text{Bin}(n, p)$ counts the number of successes over multiple (n) independent repeated trials, each with the same success probability p. It can be regarded as the sum of n i.i.d. Bernoulli variable with parameter p.

Q5. Two teams, A and B, are playing a match that consists of a series of games. Each game is won by Team A with probability p and by Team B with probability $q = 1 - p$, independently. They will keep playing until either team has won r games, where r is a fixed positive integer (so the first team to win r games wins the match). Let N be the number of games played. Find the PMF of N.

When $N = n$, there could only be two cases: (1) team A already wins whatever $r-1$ of the previous $n-1$ games, and is about to win the nth one; (2) team B already wins whatever $r-1$ of the previous $n-1$ games, and is about to win the nth one. Therefore, by LOTP,

$$P(N = n) = \binom{n-1}{r-1} p^{r-1} q^{n-r} p + \binom{n-1}{r-1} q^{r-1} p^{n-r} q = \binom{n-1}{r-1} (p^r q^{n-r} + q^r p^{n-r})$$

The support is $n = r, r+1, \ldots, 2r-1$.

We can check: when $r = 1$, obviously the winner will be decided in just one game. This accords with our PMF, where $P(N = 1) = p + q = 1$.

When $r = 2$, our PMF is

$$P(N = n) = (n-1)(p^2 q^{n-2} + q^2 p^{n-2}), \quad n = 2, 3$$

Winner can only decided in 2 or 3 games. If winner is decided in 2 games, then it must be either team winning twice, so the probability is $p^2 + q^2$, which accords with $P(N = 2)$. If winner is decided in the third game, then the two teams must draw in the first two games, and there are two ways. The third game can be won by either team. So the probability is $2pq(p + q)$, which accords with $P(N = 3) = 2(p^2 q + q^2 p)$.