反馈与讨论

2014/5/15

Important Concepts, Formulas, and Theorems

1. *Conditional probability*. The *conditional probability* of E given F, denoted by $P(E|F)$ and read as "the probability of E given F," is defined by

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

when $P(F) \neq 0$.

2. *Bayes' Theorem*. The relationship between $P(E|F)$ and $P(F|E)$ is

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)}.$$

3. *Independent*. We say E is *independent* of F if $P(E|F) = P(E)$.
4. *Product principle for independent probabilities*. The product principle for independent probabilities (Theorem 5.5) gives another test for independence. Suppose E and F are events in a sample space. Then E is independent of F if and only if $P(E \cap F) = P(E)P(F)$.
5. *Symmetry of independence*. The event E is independent of the event F if and only if F is independent of E.

6. *Independent trials process.* A process that occurs in stages is called an *independent trials process* if, for each sequence $a_1, a_2, \ldots, a_n$ with $a_i \in S_i$,

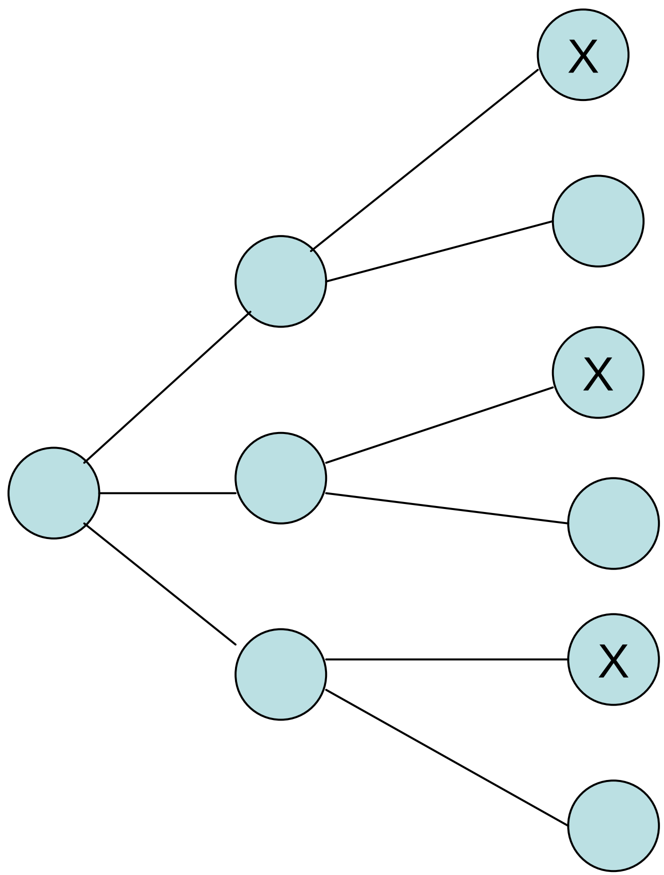
$$P(x_i = a_i | x_1 = a_1, \ldots, x_{i-1} = a_{i-1}) = P(x_i = a_i).$$

7. *Probabilities of outcomes in independent trials.* In an independent trials process, the probability of a sequence $a_1, a_2, \ldots, a_n$ of outcomes is $P(\{a_1\})P(\{a_2\}) \cdots P(\{a_n\})$.
8. *Coin flipping.* Repeatedly flipping a coin is an independent trials process.
9. *Hashing.* Hashing a list of n keys into k slots is an independent trials process with n stages.
10. *Tree diagram.* In a tree diagram for a multistage process, each level of the tree corresponds to one stage of the process. Each vertex is labeled with one of the possible outcomes at the stage it represents. Each edge is labeled with a conditional probability—the probability of getting the outcome at its right end given the sequence of outcomes that have occurred so far. Each path from the root to a leaf represents a sequence of outcomes and is labeled with the product of the probabilities along that path. This is the probability of that sequence of outcomes.

3. In three flips of a coin, is the event of getting at most one tail independent of the event that not all flips are identical?

4. What is the sample space that you use for rolling two dice, a first one and then a second one? Using this sample space, explain why the event "i dots are on top of the first die" and the event "j dots are on top of the second die" are independent if you roll two dice.

Let’s make a deal.



Important Concepts, Formulas, and Theorems

1. *Random variable.* A *random variable* for an experiment with a sample space S is a function that assigns a number to each element of S.
2. *Bernoulli trials process.* An independent trials process with two outcomes, success and failure, at each stage and probability p of success and $1 - p$ of failure at each stage is called a *Bernoulli trials process.*
3. *Probability of a sequence of Bernoulli trials.* In n Bernoulli trials with probability p of success, the probability of a given sequence of k successes and $n - k$ failures is $p^k(1 - p)^{n-k}$.
4. *The probability of* k, *successes in* n, *Bernoulli trials.* The probability of having exactly k successes in a sequence of n independent trials with two outcomes and probability p of success on each trial is given by

$$P(\text{exactly } k \text{ successes}) = \binom{n}{k} p^k(1 - p)^{n-k}.$$

5. *Binomial probability distribution.* The probabilities of k successes in n Bernoulli trials, $\binom{n}{k} p^k(1 - p)^{n-k}$, are called *binomial probabilities,* or the *binomial probability distribution.*
6. *Generating function.* The generating function for the sequence $a_0, a_1, a_2, \ldots, a_n$ is

$$\sum_{i=1}^{n} a_i x^i,$$

and the generating function for an infinite sequence $a_0, a_1, a_2, \ldots, a_n, \ldots$ is the infinite series

$$\sum_{i=1}^{\infty} a_i x^i.$$

The polynomial $(px + 1 - p)^n$ is the generating function for the binomial probabilities for n Bernoulli trials with probability p of success.
7. *Distribution function.* The function that assigns $P(X = x_i)$ to the event $X = x_i$ is called the *distribution function* of the random variable X.

8. *Expected value.* The *expected value,* or *expectation,* of a random variable X, whose values are the set $\{x_1, x_2, \ldots, x_k\}$, is defined by

$$E(X) = \sum_{i=1}^{k} x_i P(X = x_i).$$

9. *Another formula for expected values.* If a random variable X is defined on a (finite) sample space S, then its expected value is given by

$$E(X) = \sum_{s:s \in S} X(s)P(s).$$

10. *Expected value of a sum.* Suppose X and Y are random variables on the (finite) sample space S. Then

$$E(X + Y) = E(X) + E(Y).$$

    This is called the *additivity of expectation.*

11. *Expected value of a numerical multiple.* Suppose X is a random variable on a sample space S. Then $E(cX) = cE(X)$ for any number c. This result and the additivity of expectation are called the *linearity of expectation*.
12. *Expected number of successes in Bernoulli trials.* In a Bernoulli trials process, the expected number of successes is np.
13. *Indicator random variables.* A random variable that is 1 if a certain event happens and 0 otherwise is called an *indicator random variable*.
14. *Expected number of trials until success.* Suppose we have a sequence of trials in which each trial has two outcomes (success and failure) and in which the probability of success at each step is p. Then the expected number of trials until the first success is $1/p$.
15. *Geometric distribution.* The probability distribution given by $P(F^i S) = (1 - p)^i p$ is called a *geometric distribution*.

17. Evaluate the sum

$$\sum_{i=0}^{10} i \binom{10}{i} (.9)^i (.1)^{10-i},$$

which arose in computing the expected number of right answers a person would have on a 10-question test with probability .9 of answering each question correctly. First, use the binomial theorem and calculus to show that

$$10(.1 + x)^9 = \sum_{i=0}^{10} i \binom{10}{i} (.1)^{10-i} x^{i-1}.$$

Substituting $x = .9$ almost gives the sum you want on the right side of the equation, except that in every term of the sum, the power on .9 is one too small. Use some simple algebra to fix this and then explain why the expected number of right answers is 9.

Suppose we have a sequence of trials in which each trial has two outcomes, success and failure, and in which the probability of success at each step is p and $p > 0$. Then the expected number of trials until the first success is $1/p$.

Proof We consider the random variable X, which is i if the first success is on Trial i. (In other words, $X(F^{i-1}S) = i$.) The probability that the first success is on Trial i is $(1-p)^{i-1}p$, because for this to happen, there must be $i-1$ failures followed by one success. The expected number of trials is the expected value of X, which is, by the definition of expected value and the previous two sentences,

$$\begin{aligned} E(\text{number of trials}) &= \sum_{i=0}^{\infty} p(1-p)^{i-1} i \\ &= p \sum_{i=0}^{\infty} (1-p)^{i-1} i \\ &= \frac{p}{1-p} \sum_{i=0}^{\infty} (1-p)^{i} i \\ &= \frac{p}{1-p} \frac{1-p}{p^2} \\ &= \frac{1}{p}. \end{aligned}$$

To go from the third to the fourth line in the previous sequence of equations, we used the fact that

$$\sum_{j=0}^{\infty} j x^j = \frac{x}{(1-x)^2}, \tag{5.29}$$

which is true for x with absolute value less than 1. We proved a finite version of this equation as Theorem 4.6; the infinite version is even easier to prove.

4. In a card game, you remove the jacks, queens, kings, and aces from an ordinary deck of cards and shuffle them. You draw a card. If it is an ace, you are paid $1.00, and the game is repeated. If it is a jack, you are paid $2.00, and the game ends. If it is a queen, you are paid $3.00, and the game ends. If it is a king, you are paid $4.00, and the game ends. What is the maximum amount of money a rational person would pay to play this game?

Important Concepts, Formulas, and Theorems

1. *Histogram. Histograms* are graphs that show, for each integer value x of a random variable X, a rectangle of width 1 and centered at x whose height (and thus area) is proportional to the probability $P(X = x)$. Histograms can be drawn with nonunit-width rectangles. When you draw a rectangle with a base ranging from $x = a$ to $x = b$, the area of the rectangle is the probability that X is between a and b.
2. *Expected value of a constant.* If X is a random variable that always takes on the value c, then $E(X) = c$. In particular, $E\big(E(X)\big) = E(X)$.
3. *Variance.* The *variance* $V(X)$ of a random variable X is defined as the expected value of $\big(X - E(X)\big)^2$. This can also be expressed as a sum over the individual elements of the sample space S, which gives $V(X) = E\left(\big(X - E(X)\big)^2\right) = \sum_{s:s\in S} P(s)\big(X(s) - E(X)\big)^2$.
4. *Independent random variables.* Random variables X and Y are *independent* when the event that X has value x is independent of the event that Y has value y, regardless of the choice of x and y.
5. *Expected product of independent random variables.* If X and Y are independent random variables on a sample space S, then $E(XY) = E(X)E(Y)$.

6. *Variance of sum of independent random variables.* If X and Y are independent random variables, then $V(X + Y) = V(X) + V(Y)$.
7. *Standard deviation.* The square root of the variance of a random variable is called the *standard deviation* of the random variable and is denoted by σ (or by $\sigma(X)$ when there is a chance for confusion as to what random variable we are discussing).
8. *Variance and standard deviation for Bernoulli trials.* In Bernoulli trials with probability p of success, the variance for one trial is $p(1 - p)$, and for n trials, it is $np(1 - p)$. The standard deviation for n trials is $\sqrt{np(1 - p)}$.
9. *Central limit theorem.* The central limit theorem says that the sum of independent random variables with the same distribution function is approximated well as follows: The probability that the sum is between a and b is

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} \, dx$$

when the number of random variables being added is sufficiently large. This implies that the probability that a sum of independent random variables is within 1, 2, or 3 standard deviations of its expected value is approximately .68, .955, and .997, respectively. (The theorem holds more generally when the random variables have different distributions, provided that no one of them "dominates" the rest, or when the random variables are not independent, provided that not too many of them are very similar to others.)

1. Suppose a student who knows 60% of the material covered in a chapter of a textbook is going to take a five-question objective (each answer is either right or wrong, not multiple choice or true-false) quiz. Let X be the random variable that gives the number of questions the student answers correctly for each quiz in the sample space of all quizzes the instructor could construct. What is the expected value of the random variable $X - 3$? What is the expected value of $(X - 3)^2$? What is the variance of X?

2. In Problem 1, let X_i be the number of correct answers the student gets on Question i, that is, X_i is either 0 or 1. What is the expected value of X_i? What is the variance of X_i? How does the sum of the variances of X_1 through X_5 relate to the variance of X for Problem 1?

12. How many questions need to be on a short-answer test for you to be 95% sure that someone who knows 80% of the course material gets a grade between 75% and 85%?

16. This problem develops an important law of probability known as **Chebyshev's law**. Suppose you are given a real number $r > 0$ and you want to estimate the probability that the difference $|X(x) - E(X)|$ of a random variable from its expected value is more than r.

a. Let $S = \{x_1, x_2, \ldots, x_n\}$ be the sample space, and let $E = \{x_1, x_2, \ldots, x_k\}$ be the set of all x such that $|X(x) - E(X)| > r$. By using the formula that defines $V(X)$, show that

$$V(X) > \sum^{k} P(x_i)r^2 = P(E)r^2.$$

b. Show that the probability of $|X(x) - E(X)| \geq r$ is no more than $V(X)/r^2$. This is called Chebyshev's law.

18. This problem derives an intuitive law of probability known as the law of large numbers from Chebyshev's law. Informally, the **law of large numbers** says that if you repeat an experiment many times, the fraction of the time that an event occurs is very likely to be close to the probability of the event. The law applies to independent trials with probability p of success. It states that for any positive number s, no matter how small, you can make the probability of the number X of successes being between $np - ns$ and $np + ns$ as close to 1 as you choose by making the number n of trials large enough. For example, you can make the probability of the number of successes being within 1% (or 0.1%) of the expected number as close to 1 as you wish.

a. Show that the probability of $|X(x) - np| \geq sn$ is no more than $p(1-p)/s^2n$.

b. Explain why part "a" means you can make the probability of $X(x)$ being between $np - sn$ and $np + sn$ as close to 1 as you want by making n large.