TA Lecture 03 - Random Variables

March 27 - 28

School of Information Science and Technology,
ShanghaiTech University

上海科技大学
ShanghaiTech University

Outline

Main Contents Recap

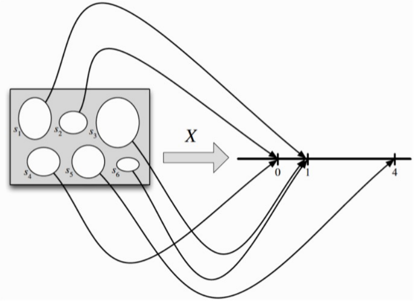
HW Problems

More Exercices

Random Variable

Definition

Given an experiment with sample space S, a *random variable* (r.v.) is a function from the sample space S to the real numbers R. It is common, but not required, to denote random variables by capital letters.



Random Variable

Definition

A random variable X is said to be *discrete* if there is a finite list of values $a_1, a_2, \ldots, a_n$ or an infinite list of values $a_1, a_2, \cdots$ such that $P(X = a_j \text{ for some } j) = 1$. If X is a discrete r.v., then the finite or countably infinite set of values x such that $P(X = x) > 0$ is called the *support* of X.

☣ **3.2.3.** In writing $P(X = x)$, we are using $X = x$ to denote an *event*, consisting of all outcomes s to which X assigns the number x. This event is also written as $\{X = x\}$; formally, $\{X = x\}$ is defined as $\{s \in S : X(s) = x\}$, but writing $\{X = x\}$ is shorter and more intuitive. Going back to Example 3.1.2, if X is the number of Heads in two fair coin tosses, then $\{X = 1\}$ consists of the sample outcomes HT and TH, which are the two outcomes to which X assigns the number 1. Since $\{HT, TH\}$ is a subset of the sample space, it is an event. So it makes sense to talk about $P(X = 1)$, or more generally, $P(X = x)$. If $\{X = x\}$ were anything other than an event, it would make no sense to calculate its probability! It does not make sense to write "$P(X)$"; we can only take the probability of an event, not of an r.v.

Independence

Definition

Random variables X and Y are said to be *independent* if

$$P(X \leq x, Y \leq y) = P(X \leq x)\, P(Y \leq y)\,,$$

for all $x, y \in \mathbb{R}$. In the discrete case, this is equivalent to the condition

$$P(X = x, Y = y) = P(X = x)\, P(Y = y)$$

for all x,y with x in the support of X and y in the support of Y.

Independence

Definition

Random variables $X_1, \ldots, X_n$ are *independent* if

$$P(X_1 \leq x_1, \cdots, X_n \leq x_n) = P(X_1 \leq x_1) \cdots P(X_n \leq x_n)$$

for all $x_1, \cdots, x_n \in \mathbb{R}$. For infinitely many r.v.s, we say that they are independent if every finite subset of the r.v.s is independent.

I.I.D.

We will often work with random variables that are independent and have the same distribution. We call such r.v.s independent and identically distributed, or i.i.d. for short.

- Independent & Identically Distributed
- Independent & NOT Identically Distributed
- Dependent & Identically Distributed
- Dependent & NOT Identically Distributed

PMF

Definition

The *probability mass function* (PMF) of a discrete r.v. X is the function p_X given by $p_X(x) = P(X = x)$. Note that this is positive if x is in the support of X, and 0 otherwise.

PMF

Theorem

Let X be a discrete r.v. with support $x_1, x_2, \ldots$ (assume these values are distinct and, for notational simplicity, that the support is countably infinite; the analogous results hold if the support is finite). The PMF p_X of X must satisfy the following two criteria:

- *Nonnegative: $p_X(x) > 0$ if $x = x_j$ for some j, and $p_X(x) = 0$ otherwise;*
- *Sums to 1: $\sum_{j=1}^{\infty} p_X(x_j) = 1$.*

CDF

Theorem

The cumulative distribution function (CDF) of an r.v. X is the function F_X given by $F_X(x) = P(X \leq x)$. When there is no risk of ambiguity, we sometimes drop the subscript and just write F (or some other letter) for a CDF.

CDF

Any CDF F has the following properties.

- Increasing: If $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.
- Right-continuous: the CDF is continuous except possibly for having some jumps. Wherever there is a jump, the CDF is continuous from the right. That is, for any a, we have

$$F(a) = \lim_{x \to a^+} F(x).$$

- Convergence to 0 and 1 in the limits:

$$\lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to \infty} F(x) = 1$$

Distribution

An experiment that can result in either a "success" or a "failure" (but not both) is called a *Bernoulli trial*. A Bernoulli random variable can be thought of as the *indicator of success* in a Bernoulli trial: it equals 1 if success occurs and 0 if failure occurs in the trial.

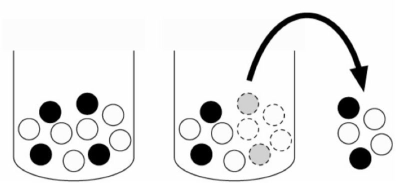
Distribution

Suppose that n *independent* Bernoulli trials are performed, each with the same success probability p. Let X be the number of successes. The distribution of X is called the *Binomial distribution* with parameters n and p. We write $X \sim \mathrm{Bin}(n, p)$ to mean that X has the Binomial distribution with parameters n and p, where n is a positive integer and $0 < p < 1$.

Distribution

An urn is filled with w white and b black balls, then drawing n balls out of the urn

- with replacement: $Bin(n, w/(w+b))$ distribution for the number of white balls obtained
- without replacement: Hypergeometric distribution



Distribution

Let C be a finite, nonempty set of numbers. Choose one of these numbers uniformly at random (i.e., all values in C are equally likely). Call the chosen number X. Then X is said to have the *Discrete Uniform distribution* with parameter C; we denote this by $X \sim \mathrm{DUnif}(C)$.

Random Variable: Geometric

Theorem

Suppose for any positive integer n, discrete random variable X satisfies

$$P(X \geq n + k | X \geq k) = P(X \geq n)$$

for $k = 0, 1, 2, \ldots$, then $X \sim \mathrm{Geom}(p)$.

Random Variable: First Success

Definition

In a sequence of independent Bernoulli trials with success probability p, let Y be the number of trials until the first successful trial, including the success. Then Y has the First Success distribution with parameter p; we denote this by $Y \sim \mathrm{FS}(p)$.

Random Variable: Negative Binomial

In a sequence of independent Bernoulli trials with success probability p, if X is the number of failures before the r^{th} success, then X is said to have the Negative Binomial distribution with parameters r and p, denoted $X \sim NBin(r, p)$.

Random Variable: Negative Binomial

Theorem

Let $X \sim \mathrm{NBin}(r, p)$*, viewed as the number of failures before the* r*th success in a sequence of independent Bernoulli trials with success probability* p*. Then we can write* $X = X_1 + \cdots + X_r$ *where the* X_i *are i.i.d.* $\mathrm{Geom}(p)$.

Random Variable: Poisson

Definition

An r.v. X has the *Poisson distribution* with parameter λ if the PMF of X is

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \ k = 0, 1, 2, \cdots$$

We write this as $X \sim \mathrm{Pois}(\lambda)$.

Random Variable: Poisson

Theorem

If $X \sim \text{Pois}(\lambda_1)$*,* $Y \sim \text{Pois}(\lambda_2)$*, and* X *is independent of* Y*, then* $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$.

Random Variable: Poisson

Let $A_1, A_2, \cdots, A_n$ be events with $p_j = P(A_j)$, where n is large, the p_j are small, and the A_j are independent or weakly dependent. Let

$$X = \sum_{j=1}^{n} I(A_j)$$

count how many of the A_j occur. Then X is approximately $\mathrm{Pois}(\lambda)$, with $\lambda = \sum_{j=1}^{n} p_j$.

Random Variable: Poisson

Theorem

If $X \sim \mathrm{Pois}(\lambda_1)$, $Y \sim \mathrm{Pois}(\lambda_2)$, *and* X *is independent of* Y, *then the conditional distribution of* X *given* $X + Y = n$ *is* $\mathrm{Bin}(n, \lambda_1/(\lambda_1 + \lambda_2))$.

Random Variable: Poisson

Theorem

If $X \sim \mathrm{Bin}(n, p)$ and we let $n \to \infty$ and $p \to 0$ such that $\lambda = np$ remains fixed, then the PMF of X converges to the $\mathrm{Pois}(\lambda)$ PMF. More generally, the same conclusion holds if $n \to \infty$ and $p \to 0$ in such a way that np converges to a constant λ.

Outline

Problem 1

Please reinterpret the following story from the Bayesian perspective.

狼来了: 从前有个放羊娃, 每天都把羊群带到山上去吃草. 山里有狼出没. 第一天, 放羊娃觉得无聊, 想要作弄山下耕作的村民. 他朝着山下大喊“狼来了!狼来了”, 村民们信以为真, 冲上山来准备帮助他, 发现被欺骗了, 大家很生气. 第二天, 放羊娃故技重施, 村民们虽然有点迟疑, 但还是冲上山来准备打狼, 结果又一次 发现被欺骗了, 大家非常生气. 第三天, 狼真的来了, 此时放羊娃慌了, 哭着向山下大喊“狼来了!狼来了!”, 请 求村民的帮助. 但这一次村民们认为他又在撒谎, 无人相信他. 最后他所有的羊都被狼吃掉了.

Problem 1 Solution

Problem 2

A fair die is rolled repeatedly, and a running total is kept (which is, at each time, the total of all the rolls up until that time). Let p_n be the probability that the running total is ever exactly n (assume the die will always be rolled enough times so that the running total will eventually exceed n, but it may or may not ever equal n).

(a) Write down a recursive equation for p_n (relating p_n to earlier terms p_k in a simple way). Your equation should be true for all positive integers n, so give a definition of p_0 and p_k for $k < 0$ so that the recursive equation is true for small values of n.

(b) Find p_7.

(c) Give an intuitive explanation for the fact that p_n â $1/3.5 = 2/7$ as $n \to \infty$.

Problem 2 Solution

Problem 3

A sequence of $n \geq 1$ independent trials is performed, where each trial ends in "success" or "failure" (but not both). Let p_i be the probability of success in the i^{th} trial, $q_i = 1 - p_i$, and $b_i = q_i - 1/2$, for $i = 1, 2, ..., n$. Let A_n be the event that the number of successful trials is even.

(a) Show that for $n = 2$, $P(A_2) = 1/2 + 2b_1 b_2$.

(b) Show by induction that $P(A_n) = 1/2 + 2^{n-1} b_1 b_2 ... b_n$ (This result is very useful in cryptography. Also, note that it implies that if n coins are flipped, then the probability of an even number of Heads is $1/2$ if and only if at least one of the coins is fair.) Hint: Group some trials into a super-trial.

(c) Check directly that the result of (b) is true in the following simple cases: $p_i = 1/2$ for some i; $p_i = 0$ for all i; $p_i = 1$ for all i.

Problem 3 Solution

Problem 4

A message is sent over a noisy channel.The message is a sequence $x_1, x_2, ..., x_n$ of n bits ($x_i \in \{0, 1\}$). Since the channel is noisy, there is a chance that any bit might be corrupted, resulting in an error ($a_0$becomes a_1 or vice versa). Assume that the error events are independent. Let p be the probability that an individual bit has an error($0 < p < 1/2$).Let $y_1, y_2, ..., y_n$ be the received message (so $y_i = x_i$ if there is no error in that bit, but $y_i = 1 - x_i$ if there is an error there).

To help detect errors, the n th bit is reserved for a parity check:x_n is defined to be 0 if $x_1 + x_2 + ... + x_{n-1}$ is even, and 1 if $x_1 + x_2 + ... + x_{n-1}$ is odd. When the message is received, the recipient checks whether y_n has the same parity as $y_1 + y_2 + ... + y_{n_1}$. If the parity is wrong, the recipient knows that at least one error occurred; otherwise, the recipient assumes that there were no errors.

Problem 4 Continued

(a) For $n = 5, p = 0.1$, what is the probability that the received message has errors which go undetected?

(b) For general n and p, write down an expression (as a sum) for the probability that the received message has errors which go undetected.

(c) Give a simplified expression, not involving a sum of a large number of terms, for the probability that the received message has errors which go undetected.

Problem 4 Solution

Problem 5

For X and Y binary digits (0 or 1), let $X \oplus Y$ be 0 if $X = Y$ and 1 if $X \neq Y$ (this operation is called exclusive or (often abbreviated to XOR), or addition mod 2).

(a) Let $X \sim \mathrm{Bern}(p)$ and $Y \sim \mathrm{Bern}(1/2)$, independently. What is the distribution of $X \oplus Y$

Problem 5 Solution

Problem 5 Continued

(b) With notation as in sub-problem(a), is $X \oplus Y$ independent of X? Is $X \oplus Y$ independent of Y? Be sure to consider both the case $p = 1/2$ and the case $p \neq 1/2$.

Problem 5 Solution

Problem 5 Continued

(c) Let $X_1, ..., X_n$ be i.i.d. (i.e., independent and identically distributed) Bern(1/2) R.V.s. For each nonempty subset J of $\{1, 2, ..., n\}$, let

$$Y_J = \oplus_{Y \in J} X_J.$$

Show that Y_J $Bern(1/2)$ and that these $2^n - 1$ R.V.s are pairwise independent, but not independent.

Problem 5 Solution

Problem 6

By LOTP for problems with recursive structure, we generate many difference equations. To solve the difference equation in the form of

$$f_{i+1} = b \cdot f_i + a \cdot f_{i-1}, i \geq 1. \tag{1}$$

where a and b are constants, we turn to the so-called characteristic equation:

$$x^2 = bx + a. \tag{2}$$

Problem 6 Continued

If such equation has two distinct roots r_1 and r_2, then the general form of f_i is

$$f_i = c \cdot r_1^i + d \cdot r_2^i, \tag{3}$$

If there is only one distinct root r, then the general form of f_i is

$$f_i = c \cdot r^i + d \cdot i \cdot r^i. \tag{4}$$

Show the mathematical principle behind the method of characteristic equation.

Problem 6 Solution

Outline

BH CH2 #62: Difference Equation

There are n types of toys, which you are collecting one by one. Each time you buy a toy, it is randomly determined which type it has, with equal probabilities. Let $p_{i,j}$ be the probability that just after you have bought your i^{th} toy, you have exactly j toy types in your collection, for $i \geq 1$ and $0 \leq j \leq n$. (This problem is in the setting of the coupon collector problem, a famous problem which we study in Example 4.3.11.)

(a) Find a recursive equation expressing p_{ij} in terms of $p_{i-1,j}$ and $p_{i-1,j-1}$, for $i \geq 2$ and $1 \leq j \leq n$.

(b) Describe how the recursion from (a) can be used to calculate $p_{i,j}$.

Solution