Probability and Statistics course

Introduction

Gabor Vigh

TTK Department of Probability Theory and Statistics
ELTE

February 18, 2026

Week 3

Materials

[Baron, 2014] [Ross, 2012]

Baron, M. (2014).
Probability and statistics for computer scientists.
CRC Press.

Ross, S. (2012).
First Course in Probability.
Pearson.

Random Variables

Frequently, when an experiment is performed, we are interested mainly in some function of the outcome as opposed to the actual outcome itself. For instance, in tossing dice, we are often interested in the sum of the two dice and are not really concerned about the separate values of each die. That is, we may be interested in knowing that the sum is 7 and may not be concerned over whether the actual outcome was $(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)$. These quantities of interest, or more formally these real-valued functions defined in the sample space, are known as random variables.

Examples

Independent trials consisting of the flipping of a coin having probability p of coming up heads are continually performed until either a head occurs or a total of n flips is made. If we let X denote the number of times the coin is flipped, then X is a random variable taking on one of the values $1, 2, 3, \ldots, n$ with respective probabilities

Random Variables

Definition

A random variable is a function $X : \Omega \to \mathbb{R}$ such that for every $t \in \mathbb{R}$

$$\{\omega \in \Omega : X(\omega) \leq t\} = \{X \leq t\}$$

is an event, that is, it is in $\mathcal{A}$. A random variable is discrete, if its range is finite or countably infinite.

(Cumulative) Distribution function

Definition ((Cumulative) Distribution Function of the Random Variable X)

$F_X(x) = P(X < x)$. Properties of the Distribution Function:

- $0 \leq F_X(x) \leq 1$
- monotonically increasing
- left-continuous
- $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.

Theorem

For any random variable X, we have $P(a \leq X < b) = F(b) - F(a)$;
$P(a < X \leq b) = F(b^+) - F(a^+)$.

Discrete distributions

Definition (Discrete Random Variable)

Its range of values is at most countably infinite, that is, it consists of elements $\{x_1, \dots, x_n, \dots\}$. Its distribution is:

$$p_i := P(X = x_i) = P(\omega : X(\omega) = x_i)$$

Definition (Expected Value of a Discrete Random Variable)

Notation: $E[X]$. Let X be a discrete random variable that takes on values $x_1, x_2, \dots$ with probabilities $p_1, p_2, \dots$. Then

$$E[X] = \sum_{k=1}^{\infty} x_k p_k, \quad \text{if the infinite sum is absolutely convergent.}$$

Variance and Standard deviation

Definition (Variance of X)

$$D^2X = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

Definition (Standard deviation of X)

$$DX = \sqrt{D^2X}$$

Bernoulli and Binomial Distribution

Definition (Bernoulli Distribution)

$P(X = k) = p^k(1-p)^k$, where $k \in \{0, 1\}$

Definition (Binomial Distribution)

$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, where $k \in \{0, 1, 2, \ldots, n\}$

Examples

Five fair coins are flipped. If the outcomes are assumed independent, find the probability mass function of the number of heads obtained.

Examples

It is known that screws produced by a certain company will be defective with probability .01, independently of each other. The company sells the screws in packages of 10 and offers a money-back guarantee if there is more than 1 screw is defective. What proportion of packages sold must the company replace?

Poisson distribution

Definition (Poisson Distribution)

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Examples

Suppose that the number of typographical errors on a single page of this book has a Poisson distribution with parameter $\lambda = \frac{1}{2}$. Calculate the probability that there is at least one error on this page.

Examples

Suppose that the probability that an item produced by a certain machine will be defective is 0.1. Find the probability that a sample of 10 items will contain at most 1 defective item.

Poisson Expected value

The expected value $E[X]$ of a discrete random variable X is defined as the sum of each possible value multiplied by its probability:

$$E[X] = \sum_{k=0}^{\infty} k \cdot P(X=k) = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

To simplify the summation, we can change the index. Let $j = k - 1$. When $k = 1$, $j = 0$. The sum becomes:

$$E[X] = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$$

The series $\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$ is the Maclaurin series expansion for the exponential function e^{λ}:

$$E[X] = \lambda e^{-\lambda}(e^{\lambda}) = \lambda e^0 = \lambda \cdot 1 = \lambda$$

Important Discrete Distributions

Name (parameters)	Values (k)	$P(X=k)$	EX	D^2X
Bernoulli (p) ($=$ Binomial ($1,\ p$))	0, 1	$p^k(1-p)^{1-k}$	p	$p(1-p)$
Binomial (n, p)	$0, 1, \dots, n$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
Poisson (λ)	$0, 1, \dots$	$\frac{\lambda^k}{k!}e^{-\lambda}$	λ	λ
Geometric/Pascal (p) ($=$ Negative binomial ($1,\ p$))	$1, 2, \dots$	$p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative binominal (n, p)	$n, n+1, \dots$	$\binom{k-1}{n-1}p^n(1-p)^{k-n}$	$\frac{n}{p}$	$\frac{n(1-p)}{p^2}$
Hyper-geometric (N, M, n)	$0, 1, \dots, n$	$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$	$n\frac{M}{N}$	$n\frac{M}{N}\left(1-\frac{M}{N}\right)\left(1-\frac{n-1}{N-1}\right)$

Homework

1. The following gambling game, known as the wheel of fortune (or chuck-a-luck), is quite popular at many carnivals and gambling casinos: A player bets on one of the numbers 1 through 6. Three dice are then rolled, and if the number bet by the player appears i times, $i = 1, 2, 3$, then the player wins i units; if the number bet by the player does not appear on any of the dice, then the player loses 1 unit. Is this game fair to the player?
2. A batch of goods contains 1% defective items. How many items must we randomly select and examine so that there is at least one defective item with a probability of at least 0.95, if we replace each selected item after inspection?
3. Roll a die as many times as the number of heads we get with two fair coins. Let X be the sum of the resulting numbers. Determine the distribution of X.
4. Flipping a coin (let p be the probability of heads), let X denote the length of the first sequence of identical outcomes. (For example, if the sequence is HHT..., then $X = 2$.) Determine the distribution of X.
5. Suppose that 3 probability practice sessions have 15, 20, and 25 students, respectively. What is the expected size of a randomly selected student's group?
6. In a beech forest, the number of beech saplings per square meter follows a Poisson distribution with a parameter $\lambda = 2.5$ saplings / m^2. What is the probability that in a 1 m^2 sample we find:
   a at most one, or
   b more than three saplings?
   c State the expected value and the standard deviation of the number of saplings.

Homework 2

1. We organize a party, and we know in advance that the number of participants is eight with probability $1/4$, nine with probability $1/3$ and ten otherwise. Calculate the expected number and the variance of the number of participants.
2. Peter is late from the university with probability 0.1 every day, independently of the other days. We know that there are 21 days in March where he goes to the university.
   a) What is the probability that he is never late in March? What is the probability of exactly 1 occasion when he is late? Of 2? In general, what is the probability that he is late in March exactly k times?
   b) What is the distribution of the number of occasions when he is late?
   c) Randomize 100 samples from the distribution of the number of late arrivals of Peter in March. Make a histogram, and calculate the average in python.
   d) What is the expectation of number of occasions when he is late?
   e) What is the variance of the number of occasions when he is late?
3. Suppose that we have 10 servers in a system. On each day, each of them breaks down with probability 0.01, independently of each other. Let Z be the number of servers (among these 10) which break down tomorrow (that is, on a given day). Calculate the probability $\mathbb{P}(Z = 2)$, and the expectation and variance of Z.

Homework 3

1. Suppose that, according to our data from the last decades, the average number of earthquakes per year is 3.42 in a given city. Suppose that the number of earthquakes in a given year has Poisson distribution, and that its expectation is equal to the average that we observed.
   a) Let us randomize a sample of size 200 from the distribution of the number of earthquakes in a given year, and make a histogram. What is the mean of the data? What is the proportion of 3 in the sample? What is the proportion of numbers which are at least 4?
   b) What is the probability that there are exactly 3 earthquakes in a year?
   c) What is the probability that there are at least 4 earthquakes in a year?
2. Suppose that the number of downloads of a webpage within an hour has Poisson distribution, and the probability that there are 0 downloads is $1/e^2$. Suppose furthermore that the number of downloads independent for disjoint time intervals. a) What is the variance of the number of downloads within an hour? b) Given that the number of downloads within an hour is at most 1, what is the probability that there are 0 downloads within this hour?