Lecture 3: Bayes theorem and discrete distributions

MVE055 / MSG810 Mathematical statistics and discrete mathematics)

Moritz Schauer

Last updated September 7, 2020, 2020

GU & Chalmers University of Technology

3 group assignment

- Skip list. Deadline Oct 7. **On Canvas.**
- Statistical research problem. Deadline Oct 16.
- Win Stone-Paper-Scissors with Markov chains. Deadline Oct. 25. Suggested programming languange: Julia

Conditional distribution

If we know some event B occurs, the probability of A given the new information B can be calculated as follows:

Conditional probability

Assume that $\mathsf{P}(B) > 0$. The conditional probability of A given B is defined as

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}. \tag{0.1}$$

Multiplication rule for probabilities

For events A and B it holds

$$\mathsf{P}(A \cap B) = \mathsf{P}(B \mid A)\mathsf{P}(A) = \mathsf{P}(A \mid B)\mathsf{P}(B).$$

The multiplication rule is useful to calculate probabilities of multiple events affecting each other.

Bayes formula

Bayes formula

For events A and B

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(B \mid A)\mathsf{P}(A)}{\mathsf{P}(B)}$$

Often it is useful to rewrite the denominator $\mathsf{P}(B)$

$$\mathsf{P}(B) = \mathsf{P}(B \mid A)\mathsf{P}(A) + \mathsf{P}(B \mid A^c)\mathsf{P}(A^c)$$

Independent events

Two events A and B are independent if knowing whether B occured does not change the probability of A

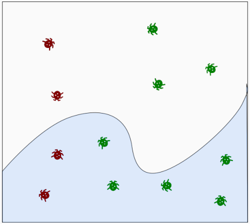
$$\mathsf{P}(A \mid B) = \mathsf{P}(A).$$

Independent events

Two events A and B are independent if $\mathsf{P}(A \cap B) = \mathsf{P}(A)\mathsf{P}(B)$.

Example with the bugs

Drawing a random bug out of the aquarium, with (g)reen and (r)ed bugs on (l)and and (w)ater.



	R	G	
L	$1/6$	$1/4$	$5/12$
W	$1/6$	$5/12$	$7/12$
	$1/3$	$2/3$	1

Frequency table and probability table

Random variables

Random variables

A random variable is a numeric quantity whose value depends on the outcome of a random event.

A random variable X is a real valued function that takes elements from Ω as argument.

We denote random variables with capital letters, often X, Y or Z.

Discrete random variables

A discrete random variables only takes a finite or countable number of values.

Integer valued random variables are automatically discrete. For now we only consider integer valued random variables.

Probability mass function

Probability mass function

For a integer valued random variable X we define the probability mass function $f(k)$ (or $f_X(k)$) by $f(k) = \mathsf{P}(X = k)$.

Probability mass function

Flip two coins... count the number of heads. Call it X.

$f(0) = \frac{1}{4}, f(1) = \frac{1}{2}$ and $f(2) = \frac{1}{4}$

Probability mass function

Not all functions are probability mass functions. Because they describe probability distributions, some conditions must hold.

$f(k)$ is a probability mass function if and only if

- $f(k) \geq 0$ for all k.
- $\sum_{k=-\infty}^{\infty} f(k) = 1$.

Distribution function

Distribution function

Assume X is a discrete random variable. Its distribution function is given by

$$F(x) = \mathsf{P}(X \leq x) = \sum_{k \leq x} f_X(k),$$

Flip two coins... count the number of heads.Call it X.
$f(0) = \frac{1}{4}, f(1) = \frac{1}{2}$ and $f(2) = \frac{1}{4}$

$$F(0) = \frac{1}{4}, \quad F(1) = \frac{1}{4} + \frac{1}{2}, \quad F(1) = 1$$

Flip two coins... count the number of heads. $f(0) = \frac{1}{4}, f(1) = \frac{1}{2}$ and $f(2) = \frac{1}{4}$

What is $P(X > 0)$?

$$P(X > 0) = f(1) + f(2) = \frac{3}{4}$$

Rule

$$\mathsf{P}(m \leq X \leq n) = \sum_{k=m}^{n} f(k)$$

for any integers m and n.

Distribution function

What is the probability to throw k times heads in a row with a fair coin?

$$f(0) = \frac{1}{2}, \quad f(1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}, \quad f(2) = \frac{1}{8}, \quad f(k) = \left(\frac{1}{2}\right)^{k+1}$$

$$P(X > 0) = f(1) + f(2) + f(3) + ... = 1 - P(X = 0) = 1 - f(0)$$

Distribution function

For $F(x)$ it holds

- $F(x)$ is increasing
- $F(x) \to 1$ for $x \to \infty$.
- $F(x) \to 0$ for $x \to -\infty$.

Distribution function

Also

- $\mathsf{P}(a < X \leq b) = F(b) - F(a)$.
- $\mathsf{P}(X > a) = 1 - F(a)$.
- For discrete $f(m) = F(m) - F(m-1)$.

Expected value

We are often interested in the "average" outcome of a random variable.

Expected value

The expected value of a random variable is defined as

$$\mathsf{E}(X) = \begin{cases} \sum\limits_{k=-\infty}^{\infty} k f(k) & \text{if } X \text{ is discrete,} \\ \int\limits_{-\infty}^{\infty} x f(x) \mathrm{d}x & \text{if } X \text{ is continuous.} \end{cases}$$

Recall: the average using fractions

Data set: grades of 24 students

$$5, 5, 6, 5, 6, 6, 6, 5, 5, 7, 6, 7, 5, 5, 5, 6, 6, 6, 5, 6, 5, 7, 6, 7$$

Table:

grade	$x_1 = 7$	$x_2 = 6$	$x_3 = 5$
fraction of students	$p_1 = 4/24$	$p_2 = 10/24$	$p_3 = 10/24$

Average One can write the average in different forms

$$\text{Average} = \frac{5+5+6+\cdots+5+7+6+7}{24}$$

$$= \frac{7\cdot 4 + 6\cdot 10 + 5\cdot 10}{24} = 7\cdot\frac{4}{24} + 6\cdot\frac{10}{24} + 5\cdot\frac{10}{24} = \sum_{i=1}^{3} x_i \cdot p_i$$

Expected value

The expected value of a discrete random variable X can also be written as

$$\mu = \mathrm{E}(X) = \sum_{i=1}^{k} x_i \cdot \underbrace{\mathrm{P}(X = x_i)}_{f(x_i)}$$

$$= x_1 \cdot \mathrm{P}(X = x_1) + x_2\ P(X = x_2) + \cdots + x_k \cdot \mathrm{P}(X = x_k)$$

Here x_i are the possible outcomes and $P(X = x_i)$ are the probabilities of each outcome.

Expected value

Flip two coins... count the number of heads.

$f(0) = \frac{1}{4}, f(1) = \frac{1}{2}$ and $f(2) = \frac{1}{4}$

$E[X] = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$

Rules for computing expected values

For the expected value,

- $\mathsf{E}(a) = a$.
- $\mathsf{E}(aX) = a\mathsf{E}(X)$.
- $\mathsf{E}(aX + b) = a\mathsf{E}(X) + b$.
- $\mathsf{E}(X + Y) = \mathsf{E}(X) + \mathsf{E}(Y)$.

Here X and Y are two random variables and a and b are constants.

Transformations

If we transform the random variables by a function h we have:

Theorem

$$\mathsf{E}(h(X)) = \sum_{k=-\infty}^{\infty} h(k) f(k)$$

Common distributions

Bernoulli distribution

The Bernoulli distribution describes a random experiment that can either succeed (with probability p) or fail (with probability $1 - p$.) Suppose we make a random experiment which succeeds with probability p and set

$$X = \begin{cases} 1, & \text{if the experiment succeeds} \\ 0, & \text{in case of failure.} \end{cases}$$

We have $f(1) = p$ and $f(0) = 1 - p$. Sometimes useful to write as $f(k) = p^k(1-p)^{1-k}$ for $k \in \{0, 1\}$.

Bernoulli distribution

A random variable X is Bernoulli distributed if it has probability mass function $f(k) = p^k(1-p)^{1-k}$, where $k = 0, 1$. We write $X \sim \mathrm{Ber}(p)$.

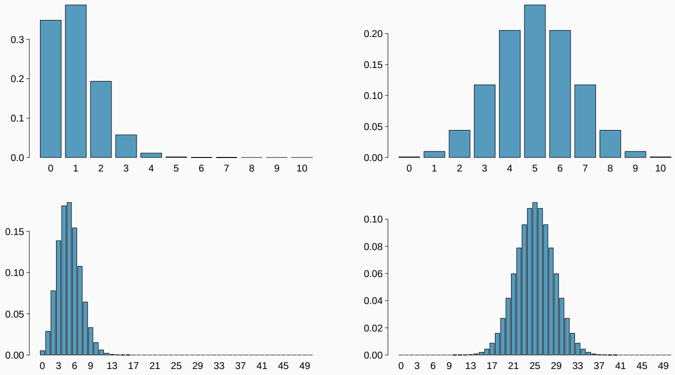
The binomial distribution

The binomial distribution describes the probability of having exactly k successes in n independent Bernoulli trials with probability of success p.

If X is Binomial with parameters n and p we write:

$$X \sim \mathrm{Bin}(n, p)$$

The binomial distribution



$n = 10$

$n = 50$

$p = 0.1$ $p = 0.5$

The binomial distribution

The binomial distribution describes the probability of having exactly k successes in n independent Bernoulli trials with probability of success p.

If X is Binomial with parameters n and p we write:

$$X \sim \mathrm{Bin}(n, p)$$

Binomial distribution

A random variable X is Binomial distributed with parameters n, p if

$$\mathrm{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Sum of binomial distributed random variables

Sum of binomial distributed random variables.

If $X_1 \sim \mathrm{Bin}(n, p)$ and $X_2 \sim \mathrm{Bin}(m, p)$ are independent, then $X_1 + X_2 \sim \mathrm{Bin}(m + n, p)$.

Geometric distribution

The experiment consists of a series of independent Bernoulli trials with probability of success equal to p.

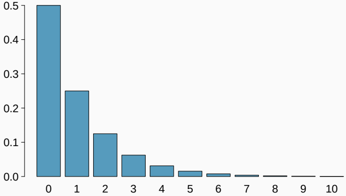
The random variable X denotes the number of trials needed to get the first success.

p is called the parameter of X.

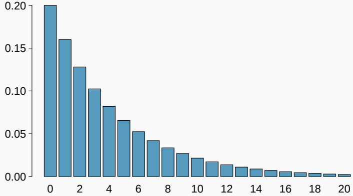
The geometric distribution

The geometric distribution describes the probability distribution of the number of failures k before the first success, for a single event succeeding with probability p.

$p = 0.5$



$p = 0.2$



The geometric distribution

Geometric distribution

A random variable X is geometrically distributed with parameters p if

$$\mathrm{P}(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \ldots$$

We write $X \sim \mathrm{Geom}(p)$.