Chapter 2: Discrete Random Variables

Mariona Segú

Assistant Professor
Thema, CY Cergy Paris Université

Fall 2023

Contents

Contents

1. Introduction to Random Variables

What is a random variable?

10 Coins

A fair coin is tossed 10 times. In this experiment, the sample space S is the set of outcomes consisting of the 2^{10} different sequences of 10 heads and/or tails that are possible. We are interested in the number of heads. We can let X stand for the real-valued function defined on S that counts the number of heads in each outcome. For example, if s is the sequence HHTTTHTTTH, then $X(s) = 4$. For each possible sequence s consisting of 10 heads and/or tails, the value $X(s)$ equals the number of heads in the sequence. The possible values for the function X are 0, 1, . . . , 10.

1. Introduction to Random Variables

Definition 1 A real-valued random variable X is any function defined in sample space S which maps the outcomes of an experiment to the real numbers

- For example, the number X of heads in the 10 tosses is a random variable. Another random variable in that example is $Y = 10 - X$, the number of tails.
- Although *genuine* randomness exists (in quantum physics for instance), in the world around us, it primarily reflects our lack of information about something.

1. Introduction to Random Variables

A person's height

Consider an experiment in which a person is selected at random from some population and her height in cm is measured. This height H is a random variable.

1. Introduction to Random Variables

Distribution of a random variable Let X be a random variable, we can determine the probabilities associated with the possible values of X.

- We treat X as a variable, i.e. we say that it can "take on" various values with the corresponding probabilities.
- In the coin flip example, where X is "coin shows tails", we just need to know that $P(X = 1) = P(X = 0) = \frac{1}{2}$.

1. Introduction to Random Variables

Tossing 10 Coins

Consider again the fair coin tossed 10 times, and let X be the number of heads. The possible values of X are $0, 1, 2, ..., 10$. For each x, $Pr(X = x)$ is the sum of the probabilities of all of the outcomes in the event X = x. Because the coin is fair, each outcome has the same probability $1/2^{10}$, and we need only count how many outcomes s have $X(s) = x$. We know that $X(s) = x$ if and only if exactly x of the 10 tosses are H. Hence, the number of outcomes s with $X(s) = x$ is the same as the number of subsets of size x (to be the heads) that can be chosen from the 10 tosses, namely

$$Pr(X = x) = \#\text{outcomes} \times \text{proba of each outcome} = \binom{10}{x}\frac{1}{2^{10}}$$

1. Introduction to Random Variables

10 Coins

For a fair coin tossed 10 times, $X =$ number of heads:

- For $X = 1$:

$$Pr(X=1) = \binom{10}{1}\frac{1}{2^{10}} = \frac{10}{1024} = 0.0098$$

- For $X = 2$:

$$Pr(X=2) = \binom{10}{2}\frac{1}{2^{10}} = \frac{45}{1024} = 0.0439$$

Contents

2. Definition and Probability Function

Definition 2 A random variable X has a **discrete** distribution if X can take on only a finite (or countably infinite) number of values $(x1, x2, ...)$.
Definition 3 If random variable X has a discrete distribution, the ***Probability Function*** **(p.f.)** of X is defined as the function

$$f(x) = P(X = x) = p(x)$$

Note: we use an uppercase letter to denote a random variable and a lowercase letter, to denote a particular value.
Note 2: we can use 3 different notations for the p.f.

2. Definition and Probability Function

- The **probability distribution** for a discrete variable X can be represented by a formula, a table, or a graph that provides $f(x) = P(X = x)$ for all x.
- Notice that $f(x) \geq 0$ for all x,
- and that any value x not explicitly assigned a positive probability is understood to be such that $f(x) = 0$.

2. Definition and Probability Function

Team selection

A supervisor in a manufacturing plant has 3 men and 3 women working for him. He wants to choose 2 workers at random. Let Y denote the number of women in his selection. Find the probability distribution for Y.

The supervisor can select 2 workers from 6 in $\binom{6}{2} = 15$ ways. Hence, there are 15 equally likely possibilities. Thus, $P(E_i) = 1/15$, for $i = 1, 2, ..., 15$. The values for Y that have nonzero probability are 0, 1, and 2. Let's compute p(0), p(1) and p(2).

2. Definition and Probability Function

Team selection

The number of ways of selecting $Y = 0$ women is $\binom{3}{0}\binom{3}{2} = 1 * 3 = 3$ because the supervisor must select zero workers from the three women and two from the three men. Thus, there are

$$p(0) = p(Y = 0) = \frac{\binom{3}{0}\binom{3}{2}}{15} = \frac{3}{15} = \frac{1}{5}$$

$$p(1) = p(Y = 1) = \frac{\binom{3}{1}\binom{3}{1}}{15} = \frac{9}{15} = \frac{3}{5}$$

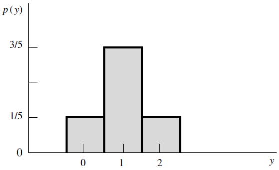
$$p(2) = p(Y = 2) = \frac{\binom{3}{2}\binom{3}{0}}{15} = \frac{3}{15} = \frac{1}{5}$$

2. Definition and Probability Function

Team selection

Table 3.1 Probability distribution for Example 3.1

y	$p(y)$
0	1/5
1	3/5
2	1/5



The most concise method of representing discrete probability distributions is with a formula

$$p(y) = p(Y = y) = \frac{\binom{3}{y}\binom{3}{2-y}}{15}$$

2. Definition and Probability Function

Some properties of discrete random variables:

1. If $\{x_1, x_2, ...\}$ is the set of all possible values of X, then for any $x \notin \{x_1, x_2, ...\}$, $f(x) = 0$.
2. $0 \leq f(x) \leq 1$
3. Sum over all possible values of x with nonzero probability

$$\sum_x f(x) = 1$$

2. Definition and Probability Function

A die

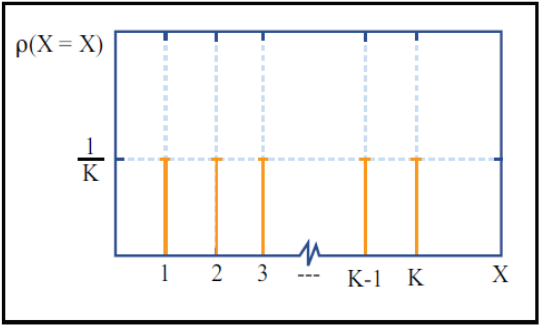
If X is the number we rolled with a die, all integers $1, 2, 3, 4, 5, 6$ are equally likely. More generally, we can define the discrete uniform distribution over the numbers $x_1, x_2, ..., x_6$ by its p.f.

$$f(x) = \begin{cases} \frac{1}{6} & \text{if } x \in \{x_1, x_2, ..., x_6\} \\ 0 & \text{otherwise} \end{cases}$$

This corresponds to the simple probabilities for an experiment with sample space $S = \{x_1, x_2, ..., x_6\}$.

2. Definition and Probability Function

Probability distribution for a die with k faces (instead of 6):



2. Definition and Probability Function

5 coins

Suppose we toss 5 fair coins independently from another and define a random variable X which is equal to the observed number of heads. How many values can X take? Find $n(S)$, $n(\text{"}k \text{ heads"})$, and $P(X = x)$

2. Definition and Probability Function

5 coins

Suppose we toss 5 fair coins independently from another and define a random variable X which is equal to the observed number of heads. Then by our counting rules, $n(S) = 2^5 = 32$, and $n(\text{"}k \text{ heads"}) = \binom{5}{k}$, using the rule on combinations. Therefore

$$P(X=0) = \binom{5}{0}\frac{1}{32} = \frac{1}{32} \qquad P(X=1) = \binom{5}{1}\frac{1}{32} = \frac{5}{32}$$

$$P(X=2) = \binom{5}{2}\frac{1}{32} = \frac{10}{32} \qquad P(X=3) = \binom{5}{3}\frac{1}{32} = \frac{10}{32}$$

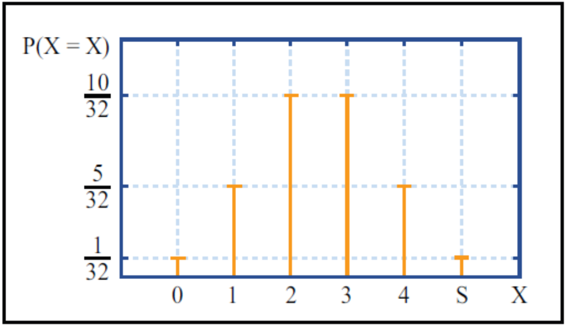
$$P(X=4) = \binom{5}{4}\frac{1}{32} = \frac{5}{32} \qquad P(X=5) = \binom{5}{5}\frac{1}{32} = \frac{1}{32}$$

2. Definition and Probability Function

5 coins

Suppose we toss 5 fair coins independently from another and define a random variable X which is equal to the observed number of heads.
Draw the PF in a graph.

2. Definition and Probability Function



2. Definition and Probability Function

- Note that in the die roll example, every single outcome corresponded to exactly one value of the random variable.
- In contrast for the five coin tosses there was a big difference in the number of outcomes corresponding to $X = 2$ compared to $X = 0$.
- So mapping outcomes into realizations of a random variable may lead to highly skewed distributions even though the underlying outcomes of the random experiment may all be equally likely.

Wooclap

Question #11 and #12

Contents

3. The Expected Value

The Expected Value

- So far, we have seen the probability distribution which is a model of the *real* empirical distribution of a variable
- We now look at descriptive measures of such distribution
- We can summarize the most important characteristics without having to give the entire density function

3. The Expected Value

Definition Let Y be a discrete random variable with the probability function $p(y)$. Then the expected value of Y, $E(Y)$, is defined by

$$E(Y) = \sum_y yp(y)$$

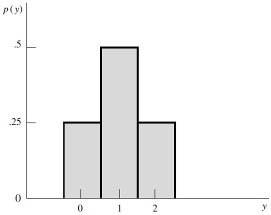
In words: *"How often is each value repeated?"*
If $p(y)$ is an accurate characterization of the population frequency distribution, then $E(Y) = \mu = \frac{\sum_{i=1}^n y_i}{n}$, the population mean.
(Recall Mean definition from Ch1 Intro Go)

3. The Expected Value

Expected Value

Table 3.2 Probability distribution for Y

y	$p(y)$
0	1/4
1	1/2
2	1/4



If we run the experiment 4M times, we should expect to obtain:

- 1M times y=0
- 2M times y=1
- 1M times y=2

$$\mu = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{1M\times 0 + 2M\times 1 + 1M\times 2}{4M} = 0\times 1/4 + 1\times 1/2 + 2\times 1/4 = \sum_y yp(y) = 1$$

3. The Expected Value

The formula of the expected value can be applied to a **function of a random variable** (not only the random variable itself).

- If g(Y) is a function of random variable Y, then, its expected value is

$$E[g(Y)] = \sum_y g(y)p(y)$$

3. The Expected Value

We are looking at numerical descriptive measures of $p(y)$. $E(y)$ provides the mean of Y. Let's now look at the variance and standard deviation:
(Recall Var definition from Ch1 Intro Go)

- **Variance**: we need to find the mean of the function $g(Y) = (Y - \mu)^2$

$$V(Y) = E(g(Y)) = E[(Y - \mu)^2] = \sum_y (y - \mu)^2 p(y)$$

- The **Standard deviation** is $\sqrt{V(Y)}$.
- As before, if $p(y)$ accurately describes the population, then $V(Y) = \sigma^2$ and the standard deviation is σ.

3. The Expected Value

Example

Find the mean, the var and the std. dev.

Table 3.3 Probability distribution for Y

y	$p(y)$
0	1/8
1	1/4
2	3/8
3	1/4

3. The Expected Value

Example

The mean

$$\mu = E(Y) = \sum_{y=0}^{3} yp(y)$$

$$= 0 \times 1/8 + 1 \times 1/4 + 2 \times 3/8 + 3 \times 1/4 = 1.75$$

3. The Expected Value

Example

The Variance

$$\sigma^2 = E[(Y-\mu)^2] = \sum_{y=0}^{3}(y-\mu)^2 p(y)$$

$$= (0-1.75)^2(1/8)+(1-1.75)^2(1/4)+(2-1.75)^2(3/8)+(3-1.75)^2(1/4)$$
$$= 0.9375$$

The standard deviation

$$\sigma = \sqrt{\sigma^2} = \sqrt{0.9375} = 0.97$$

3. The Expected Value

Properties of the Expected Value

1. The expected value of a constant c is $E[c] = c$
2. If $g(Y)$ is a function of Y and c is a constant

$$E[cg(Y)] = cE[g(Y)]$$

3. Let $g_1(Y), g_2(Y), ..., g_k(Y)$ be functions of Y, then

$$E[g_1(Y)+g_2(Y)+...+g_k(Y)] = E[g_1(Y)]+E[g_2(Y)]+...+E[g_k(Y)]$$

3. The Expected Value

We can use these properties to find that

$$V(Y) = \sigma^2 = E[(Y-\mu)^2] = E[Y^2] - \mu^2$$

Proof:

$$\begin{aligned}\sigma^2 = E[(Y-\mu)^2] &= E[(Y^2 - 2\mu Y + \mu^2)] \\ &= E[Y^2] - E[2\mu Y] + E[\mu^2] \\ &= E[Y^2] - 2\mu E[Y] + \mu^2 \\ &= E[Y^2] - 2\mu^2 + \mu^2 = E[Y^2] - \mu^2\end{aligned}$$

since $\mu = E[Y]$

3. The Expected Value

Example

Use the second definition of $V(Y)$ to find the variance in the previous example. To use $V(Y) = E[Y^2] - \mu^2$, we first need to find $E[Y^2]$ using $E[g(Y)] = \sum_y g(y)p(y)$:

$$E[Y^2] = \sum_y y^2 p(y) = 0^2 \times (1/8) + 1^2 \times (1/4) + 2^2 \times (3/8) + 3^2 \times (1/4) = 4$$

Hence, using the last formula:

$$\sigma^2 = E[Y^2] - \mu^2 = 4 - 1.75^2 = 0.9375$$

Wooclap

Question #13 and #14

Contents

5. Cumulative Distribution Function

So far, we have seen the p.f. of a discrete and a random variable

$$f(x) = P(X = x)$$

A p.f. reports the probability that a given variable X equals one particular value x.

Now, let's look at the **Cumulative Distribution Function (c.d.f.)**

5. Cumulative Distribution Function

The **cumulative distribution function (c.d.f.)** F of a random variable X is defined for each real number as

$$F(x) = P(X \leq x)$$

- In words: the c.d.f. at x gives the probability that the random variable takes a value ***less than or equal to*** x instead of *exactly equal* like the p.f.
- Note: since X is discrete, note that $P(X \leq x)$ may be different from $P(X < x)$.
- In the definition of the c.d.f., we'll always use X "less or equal to" x.

5. Cumulative Distribution Function

Since the c.d.f. is a probability, it inherits all the properties of probability functions:

- Property 1 The c.d.f. only takes values between 0 and 1

$$0 \leq F(x) \leq 1 \qquad \text{for all } x \in \mathbb{R}$$

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

$$\lim_{x \to \infty} F(x) = 1$$

If we let $x \to -\infty$, the event $(X \leq x)$ becomes "close" to the impossible event in terms of its probability of occurring, whereas if $x \to \infty$, the event $(X \leq x)$ becomes almost certain

- Property 2 F is monotonic, non decreasing in x, i.e.

$$F(x1) \leq F(x2) \qquad \text{for } x1 < x2$$

5. Cumulative Distribution Function

CDF of a Random Variable

Suppose that Y has distribution $p(y)$ with $y = 0, 1, 2$. Find $F(y)$ if the PF is:

$$p(y) = \binom{2}{y} \frac{1}{2}^{y} \frac{1}{2}^{2-y}$$

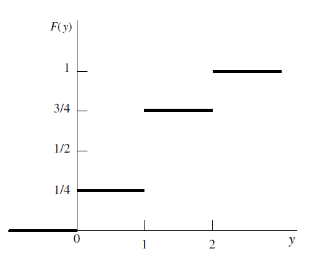
which yields $p(0) = 1/4$, $p(1) = 1/2$ and $p(2) = 1/4$

5. Cumulative Distribution Function

CDF of a Random Variable

Then,

$$F(y) = P(Y \leq y) = \begin{cases} 0 & \text{for } y < 0 \\ \frac{1}{4} & \text{for } 0 \leq y < 1 \\ \frac{3}{4} & \text{for } 1 \leq y < 2 \\ 1 & \text{for } y \geq 2 \end{cases}$$



5. Cumulative Distribution Function: Continuity

- A cumulative distribution function (c.d.f.) does not have to be continuous.
- For a **discrete random variable**, the c.d.f. is a **step function** with jumps at the values where the random variable has positive probability.
- To describe this formally, we use the notions of **left-limit** and **right-limit** at a point x.
- $F(x^-)$ denote the limit of the values of $F(y)$ as y approaches x from the left, that is, through values smaller than x.

5. Cumulative Distribution Function: Limits

- The **left-limit** of the c.d.f. at x is

$$F(x^-) = \lim_{y \to x,\, y < x} F(y)$$

- The **right-limit** of the c.d.f. at x is

$$F(x^+) = \lim_{y \to x,\, y > x} F(y)$$

- A c.d.f. has a **jump discontinuity** at x whenever

$$F(x^-) \neq F(x^+)$$

- For discrete random variables, the size of the jump at x is exactly $P(X = x)$.

5. Cumulative Distribution Function

A die

Consider again the experiment of rolling a die, where the random variable X corresponds to the number we rolled. Then the c.d.f. of X is given by

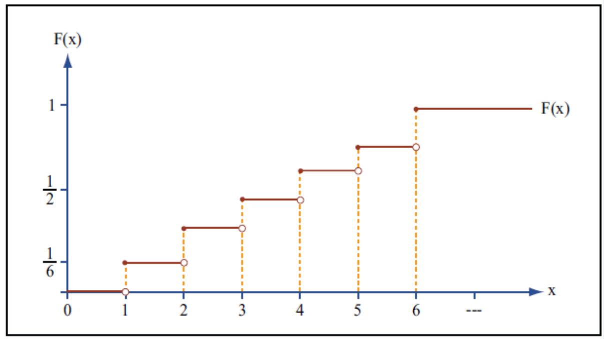
$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{1}{6} & \text{if } 1 \leq x < 2 \\ ... & ... \\ \frac{5}{6} & \text{if } 5 \leq x < 6 \\ 1 & \text{if } x \geq 6 \end{cases}$$

which has discontinuous jumps at the values $1, 2, ..., 6$.

5. Cumulative Distribution Function

A die

c.d.f. of a die roll, it is nondecreasing



5. Cumulative Distribution Function: Right-Continuity

- By definition, every c.d.f. is **continuous from the right**.
- This means that when we approach a point x from values larger than x, the c.d.f. matches exactly:

$$F(x) = F(x^+) = \lim_{y \to x,\, y > x} F(y)$$

- For a **discrete random variable**, the c.d.f. has jumps. The "filled dot" is always on the **right side of the jump**.
- In contrast, the left-limit is smaller:

$$F(x^-) < F(x) = F(x^+)$$

- The size of the jump at x equals $P(X = x)$.

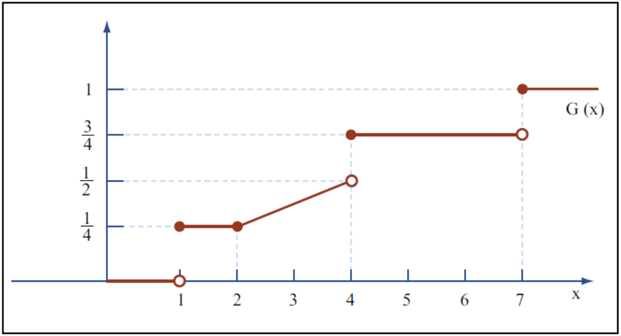
5. Cumulative Distribution Function

Let's show some more properties

- Property 3 For any given x, $P(X > x) = 1 - P(X \leq x) = 1 - F(x)$
- Property 4 For two real numbers $x_1 < x_2$, $P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1)$
- Property 5 For any x, $P(X < x) = F(x^-)$
- Property 6 For any x, $P(X = x) = F(x^+) - F(x^-)$

5. Cumulative Distribution Function

Let's check whether the function $G(x)$ in the following graph is a c.d.f.



5. Cumulative Distribution Function

Let's apply properties 3-6 to the previous graph

- $P(X > 4)$
- $P(2 < X \leq 4)$
- $P(X < 4)$
- $P(X = 4)$

5. Cumulative Distribution Function

Let's apply properties 4-7 to the previous graph

- $P(X > 4) = 1 - F(4) = 1 - \frac{3}{4} = \frac{1}{4}$
- $P(2 < X \leq 4) = F(4) - F(2) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$
- $P(X < 4) = F(4^-) = \frac{1}{2}$
- $P(X = 4) = F(4^+) - F(4^-) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$

Wooclap

Question #15 and #16

Contents

5. Discrete Random Variables: Key Examples

- So far, we defined discrete random variables and their probability functions (PFs).
- To make these ideas concrete, we now turn to two of the most important distributions:
  1. **Binomial distribution:** models the number of successes in a fixed number of independent trials.
  2. **Poisson distribution:** models the number of events occurring in a fixed interval of time or space.
- These examples are fundamental because:
  - They appear in many real-world applications.
  - They illustrate how discrete PFs are constructed.
  - They provide building blocks for more advanced probability models.

5. The Binomial Distribution

Binomial experiment: a sequence of identical and independent trials with the following properties:

1. A fixed number of trials, n.
2. Each trial has only two possible outcomes: **success (S)** or **failure (F)**.
3. The probability of success on each trial is constant: $P(S) = p$, and the probability of failure is $P(F) = q = 1 - p$.
4. The trials are independent (the outcome of one does not affect the others).
5. The random variable of interest is

$$Y = \text{number of successes in } n \text{ trials.}$$

5. The Binomial Distribution

Is this a binomial distribution?

1. An early-warning detection system for aircraft consists of four identical radar units operating independently of one another. Suppose that each has a probability of .95 of detecting an intruding aircraft. When an intruding aircraft enters the scene, the random variable of interest is Y, the number of radar units that do not detect the plane.
2. Suppose that 40% of a large population of registered voters favor candidate Jones. A random sample of $n = 10$ voters will be selected, and Y, the number favoring Jones, is to be observed. Does this experiment meet the requirements of a binomial experiment?

5. The Binomial Distribution

Is this a binomial distribution?

1. Aircraft radar
   - We need to assess the five requirement.
   - Random variable Y: number of radar units not detecting an aircraft.
   - In binomial experiments, Y denotes successes. Here, "not detecting" is a success.
     1. Four identical trials: each radar unit's detection check.
     2. Two outcomes: S (not detected) and F (detected).
     3. Equal detection probability: $p = P(S) = 0.05$.
     4. Independent trials: units operate separately.
     5. Y is number of successes in four trials.
   - Conclusion: Binomial experiment with $n = 4$, $p = 0.05$, $q = 0.95$.

5. The Binomial Distribution

Is this a binomial distribution?

2. Voters
   1. 10 random people form nearly identical trials.
   2. Outcome: favoring Jones (S) or not (F).
   3. Probability of a person favoring Jones: 0.4. Same probability across trial (since for large voter populations, removing one doesn't significantly change the fraction favoring Jones.)
   4. Such trials are approximately independent.
   5. Random variable: number of successes in 10 trials.

   Conclusion: This is an approximate binomial sampling problem.

5. The Binomial Distribution

Let's generalize,

- Consider n independent trials, each with probability of success p and probability of failure $1-p$.
- If we fix one particular sequence with exactly y successes and $n-y$ failures (for example: S–S–F–S–F–...), the probability of this sequence is
$$p^y(1-p)^{n-y}.$$
- This comes directly from the independence assumption: we multiply the probability of each outcome.

5. The Binomial Distribution

We are not interested in the *order* of successes and failures, only in the *total number of successes*.

- The number of distinct sequences with exactly y successes among n trials is
$$\binom{n}{y}.$$
- Therefore, the total probability of observing y successes is
$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}.$$
- This is what we call a **binomial distribution**

5. The Binomial Distribution

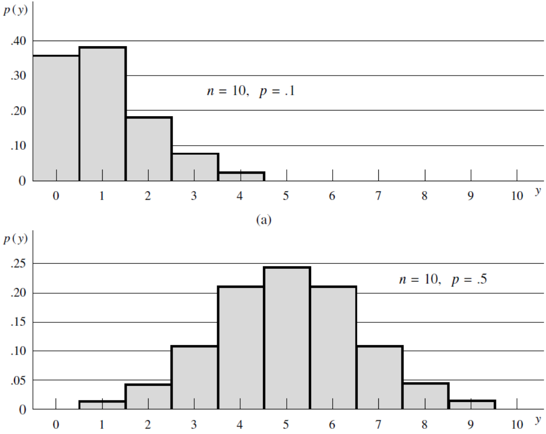
Definition 4 A random variable X with **Probability Function**

$$f(y) = P(Y = y) = \begin{cases} \binom{n}{y} p^y (1-p)^{n-y} & \text{if } x \in \{1, 2, ..., n\} \\ 0 & \text{otherwise} \end{cases}$$

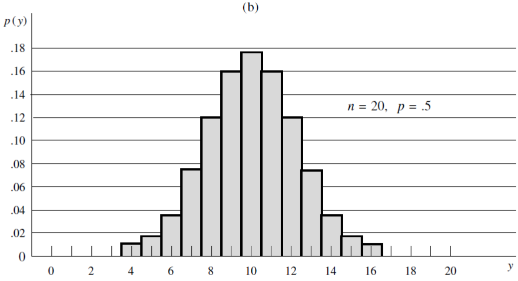
is said to follow a **binomial distribution** with parameters p and n, we write it as

$$Y \sim B(n, p)$$

5. The Binomial Distribution



5. The Binomial Distribution



(c)

5. The Binomial Distribution

The cumulative density function of a binomial is:

$$F(y) = P(Y \leq y) = \sum_{i=0}^{y} \binom{n}{y} p^y (1-p)^{n-y}$$

Computing this by hand can be very tedious... Table 1 in Tables Appendix provide the binomial tabulation for some values of n, in the form $\sum_{y=0}^{a} p(y)$.

5. The Binomial Distribution

Tricking your classmates

In order to make some money off your classmates, you obtained a rigged 1 euro coin that comes up heads with a probability of $p_R = 4/5$. Unfortunately, that coin got mixed up with your regular small coins, and only after you spent 8 out of 9 coins you notice your mistake. You toss the coin 10 times, and it comes up heads for a total of 8 times. Would it be a good idea to continue to rip off your friends or are you now stuck with a regular (fair) coin with $p_F = 1/2$? I.e. what is $P(A|B)$ for A ="remaining coin is bent" and B ="8 heads out of 10"?

5. The Binomial Distribution

Tricking your classmates

If the coin is fair,

$$P(B|A^C) = \binom{10}{8} p_F^8 (1-p_F)^{10-8} = \binom{10}{8} \frac{1}{2^{10}}$$

If it is rigged,

$$P(B|A) = \binom{10}{8} p_R^8 (1-p_R)^{10-8} = \binom{10}{8} \frac{4^8 \cdot 1^2}{5^{10}}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} =$$

$$\frac{\binom{10}{8} \frac{4^8 \cdot 1^2}{5^{10}} \frac{1}{9}}{\binom{10}{8} \frac{4^8 \cdot 1^2}{5^{10}} \frac{1}{9} + \binom{10}{8} \frac{1}{2^{10}} \frac{8}{9}} = \frac{\frac{4^8 \cdot 1^2}{5^{10}}}{\frac{4^8 \cdot 1^2}{5^{10}} + \frac{8}{2^{10}}} = 46.21\%$$

5. The Binomial Distribution

Tricking your classmates

Still, the probability of heads is quite high...

$$P(H|B) = p_R \cdot P(A|B) + p_F \cdot P(A^C|B) =$$
$$\frac{4}{5} \cdot 46.21\% + \frac{1}{2} \cdot 53.79\% = 63.86\%$$

Here, we are using the formula of the total probability (adding the conditionality to B):

$$P(H) = P(H|A)P(A) + P(H|A^C)P(A^C) = p_R \cdot P(A) + p_F \cdot P(A^C)$$

5. The Binomial Distribution

The Expected Value of a Binomial distribution

- Let Y be a binomial random variable with n trial and a success probability p, then

$$\mu = E[Y] = np$$

and

$$\sigma^2 = V(Y) = np(1-p)$$

(we will not show the proof in this class)

Wooclap

Question #17 and #18

Contents

6. The Poisson Distribution

The Poisson probability distribution provides a good model for the probability distribution of the number Y of rare events that occur in space, time, volume, or any other dimension, where λ is the average value of Y. Examples:

- Number of automobile accidents (or other type of accidents)
- Number of typing errors
- Number of selfie-induced deaths
- Likelihood of encountering a wolf in the mountains

6. The Poisson Distribution

A random variable has a Poisson distribution if its PF has the following shape:

$$p(y) = \frac{\lambda^y}{y!} e^{-\lambda}$$

for $y = 0, 1, 2...$ and $\lambda > 0$

6. The Poisson Distribution

The cdf of a Poisson is:

$$F(Y \leq y) = \sum_{i=0}^{y} \frac{\lambda^y}{y!} e^{-\lambda}$$

for $y = 0, 1, 2...$ and $\lambda > 0$

Table 3 in Tables Appendix provide the Poisson tabulation for some values of λ.

6. The Poisson Distribution

Example

Suppose that a random system of police patrol is devised so that a patrol officer may visit a given beat location $Y = 0, 1, 2, 3, \ldots$ times per half-hour period, with each location being visited an average of once per time period. Assume that Y possesses, approximately, a Poisson probability distribution.

1. Calculate the probability that the patrol officer will miss a given location during a half-hour period.
2. What is the probability that it will be visited once?
3. Twice?
4. At least once?

6. The Poisson Distribution

Example

1. Calculate the probability that the patrol officer will miss a given location during a half-hour period.
   For this example the time period is a half-hour, and the mean number of visits per half-hour interval is $\lambda = 1$. Then

$$p(y) = \frac{(1)^y e^{-1}}{y!} = \frac{e^{-1}}{y!}$$

   The probability that the location is missed corresponds to $y = 0$

$$p(0) = \frac{e^{-1}}{0!} = e^{-1} = 0.368$$

6. The Poisson Distribution

Example

2. What is the probability that it will be visited once?

$$p(1) = \frac{e^{-1}}{1!} = e^{-1} = 0.368$$

3. Twice?

$$p(2) = \frac{e^{-1}}{2!} = \frac{e^{-1}}{2} = 0.184$$

4. At least once?

$$P(Y \geq 1) = 1 - P(Y < 1) = 1 - p(0) = 1 - 0.368 = 0.632$$

6. The Poisson Distribution

Poisson is useful because it can be used to approximate the binomial probability function for large n and small p.
In this case, $\lambda = np$: Approximation is good when λ is roughly less than 7.

Binomial Approximation

Suppose that Y possesses a binomial distribution with $n = 20$ and $p = .1$.

1. Find the exact value of $P(Y \leq 3)$ using the table of binomial probabilities, Table 1, Appendix.
2. Use Table 3, Appendix 3, to approximate this probability, using the Poisson distribution. Compare the exact and approximate values for $P(Y \leq 3)$.

6. The Poisson Distribution

Binomial Approximation

1. the exact (accurate to three decimal places) value of $P(Y \leq 3) = .867$.
2. If W is a Poisson-distributed random variable with $\lambda = np = 20(.1) = 2$, $P(Y \leq 3)$ is approximately equal to $P(W \leq 3)$. Table 3, gives $P(W \leq 3) = .857$.

6. The Poisson Distribution

The expected value of a Poisson distribution is

$$\mu = E(Y) = \lambda$$

The variance is

$$\sigma^2 = V(Y) = \lambda$$

Other discrete distributions

- The Geometric Probability distribution
- The Negative Binomial
- The Hypergeometric distribution

Do we have time for any of those?

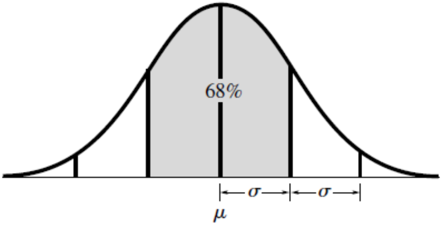
Wooclap

Question #19 and #20

Contents

Empirical Rule for Normal Distributions

- Real-life data often have mound-shaped distributions.
- This can be approximated by a bell-shaped curve called a normal distribution.
- Data with a bell-shaped form follow the following rule
- **Empirical Rule:**
  - $\mu \pm \sigma$: Contains $\approx$ 68% of measurements.
  - $\mu \pm 2\sigma$: Contains $\approx$ 95% of measurements.
  - $\mu \pm 3\sigma$: Contains almost all measurements.



Tchebysheff's Theorem

- In certain scenarios, empirical rule may not provide useful approximations.
- Tchebysheff's theorem offers a lower bound for the probability of Y being within an interval $\mu \pm k\sigma$.
- **Tchebysheff's theorem**

$$P(|Y-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$
$$P(|Y-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- The theorem:
  - Is valid for any probability distribution.
  - Provides conservative estimates.
  - Doesn't contradict empirical rule (verify!).

Tchebysheff's Theorem

Customer counter

The number of customers per day at a sales counter, Y, has been observed for a long period of time and found to have mean 20 and standard deviation 2. The probability distribution of Y is not known. What can be said about the probability that, tomorrow, Y will be greater than 16 but less than 24?

Tchebysheff's Theorem

Customer counter

- Tchebysheff's Theorem:

$$P(|Y-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

- Plugging in given values ($\mu = 20, \sigma = 2$) with $k = 2$:

$$P(16 < Y < 24) = P(\mu - 2\sigma < Y < \mu + 2\sigma) \geq \frac{3}{4}$$

- High likelihood of tomorrow's customer count being between 16 and 24.

Introduction

Statistics - characterizing a measurement

Numerical descriptive measures

1. Measure of central tendency: the mean.

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$\bar{y}$ is the sample mean, μ is the population mean

Two measurement can have the same mean but a very different distribution.

Back

Introduction

Statistics - characterizing a measurement

Numerical descriptive measures

2. Measures of dispersion:
   - The variance: the *average* deviation from the mean

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

   - The standard deviation

$$s = \sqrt{s^2}$$

The population parameters are noted σ^2 and σ, respectively

Back

Wooclap

Question #21 and #22