Probability and Statistics Notes

Chapter Two

Jesse Crawford

Department of Mathematics
Tarleton State University

Outline

Random Variables

Definition

If A and B are sets, and f is a function from A to B, we write

$$f : A \to B.$$

Definition

A *random variable* X is a function from the sample space S to the real numbers.

$$X : S \to \mathbb{R}$$

Example

Two coins are flipped and the resulting sequence of heads/tails is noted. Let X be the number of heads in the sequence.

Example

Assuming the coins are fair and independent, calculate $P(X = 1)$ and $P(X \geq 1)$.

- $(X = 1)$ is shorthand for $\{s \in S \mid X(s) = 1\} = \{HT, TH\}$
- $(X \geq 1)$ is shorthand for $\{s \in S \mid X(s) \geq 1\} = \{HT, TH, HH\}$

Definition

If $A \subseteq \mathbb{R}$, define the event

$$(X \in A) = \{s \in S \mid X(s) \in A\}.$$

Example

Roll two independent fair dice, and let X be the sum of the rolls. Calculate $P(X = x)$, for $x = 5, 6, 7, 8, 9$.

Definition

The *support* of a random variable X is the set of possible values of X,

$$\text{supp}(X) = \{X(s) \mid s \in S\}$$

Definition

A random variable is called *discrete* if its support is countable (is finite or can be put in one-to-one correspondence with the positive integers).

Example

A fair coin is flipped until the result is heads, and X is the number of flips that occur.

- What is the support of X?
- Is X a discrete random variable?

Definition

The *probability mass function* f of a discrete random variable X is

$$f : \mathbb{R} \to [0, 1]$$

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

- abbreviated p.m.f.
- also called the probability *distribution* function or probability *density* function (p.d.f.)

Proposition

A function $f : \mathbb{R} \to [0, 1]$ is the p.m.f. of some random variable if and only if

- $f(x) \geq 0$, for $x \in \mathbb{R}$, and
- $$\sum_{x \in \mathbb{R}} f(x) = 1.$$

Example

Let X be a random variable with p.m.f.

$$f(x) = \begin{cases} cx^2, & x = 1, 2, 3, 4, 5 \\ 0 & \text{otherwise.} \end{cases}$$

Find c.

Example

Let X be the number of aces in a five-card poker hand.

- Find the p.m.f. of X.
- Draw a probability histogram for X.
- The number of aces in each of ten poker hands is listed below:

  $$0, 0, 0, 1, 0, 0, 2, 0, 1, 1$$

  Draw a relative frequency histogram for this data on the same set of axes as the probability histogram.

Definition (Hypergeometric Distribution)

Setting:

- Set of objects of two types
- $N =$ total number of objects
- $N_1 =$ number of objects of the 1st type
- $N_2 =$ number of objects of the 2nd type
- Select n objects randomly without replacement
- $X =$ number of objects in sample of 1st type
- The p.m.f. of X is

$$P(X = x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}.$$

- X is said to have a *hypergeometric distribution*.

Outline

Example

When you buy a scratch-off lottery ticket, you have an 80% chance of winning nothing, a 15% chance of winning \$2, and a 5% chance of winning \$10. If the ticket costs \$1, should you buy one?

Definition

Suppose X is a discrete random variable with p.m.f. f. Then the *expected value* of X is

$$E(X) = \sum_{x \in \mathbb{R}} x f(x),$$

assuming the series converges absolutely. Otherwise, the expected value does not exist.

Example

Let X be the number of heads occurring when a fair coin is flipped 3 times.

- What is the expected value of X?
- Find $E(X^2 + 7X)$
- Find $E(5X + 4)$

Suppose $u : \mathbb{R} \to \mathbb{R}$.

$$E(u(X)) = \sum_{x \in \mathbb{R}} u(x)f(x)$$

For random variables X and Y and a constant c,

- $E(X + Y) = E(X) + E(Y)$
- $E(cX) = cE(X)$
- $E(c) = c$

An Insurance Policy

Example

An automobile insurance policy has a deductible of \$500. Let X be the cost of damages to a vehicle in an accident, and assume X has the following p.m.f.

x	0	250	500	1000	2000
$f(x)$	0.1	0.2	0.4	0.2	0.1

If an accident occurs, what is the expected value of the payment made by the insurance company?

Expected Value for the Hypergeometric Distribution

Example

In a club with a 100 members, 60 members approve of the president. In a random sample of size 5, let X be the number of people who approve of the mayor. Find the expected value of X.

Let X have a hypergeometric distribution where

- $N_1 =$ number of objects of type 1
- $N =$ total number of objects
- $n =$sample size

$$E(X) = \frac{N_1}{N} n$$

Let X have a hypergeometric distribution where

- N_1 = number of objects of type 1
- N = total number of objects
- n =sample size

$$E(X) = \frac{N_1}{N} n$$

Example

If you have 10 red pens and 4 blue pens, and you select 6 pens at random, what is the expected value of the number of blue pens in your sample?

Outline

- $E(X)$ = expected value of X
- $E(X)$ = "average value" of X
- $E(X)$ also called the *mean* of X
- Alternative notation:

$$\mu = E(X) \text{ or } \mu_X = E(X)$$

Definition

The *variance* of X is

$$\mathrm{Var}(X) = E[(X-\mu)^2] = E(X^2) - \mu^2$$

The *standard deviation* of X is

$$\sigma = \sqrt{\mathrm{Var}(X)}.$$

Definition

The rth moment of X about b is

$$E[(X - b)^r].$$

The rth moment of X about the origin, $E(X^r)$, is usually just called the rth moment of X.

Definition

Let $x_1, x_2, \ldots, x_n$ be a sample.

- The *sample mean* is
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$
- The *sample variance* is
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$
- The sample variance can be computed more easily as follows:
$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}{n-1}.$$
- The *sample standard deviation is* $s = \sqrt{s^2}$.

The variance of a hypergeometric random variable is

$$\text{Var}(X) = np(1-p)\frac{N-n}{N-1},$$

where $p = \frac{N_1}{N}$.

Outline

Definition

- A *Bernoulli trial* is a random experiment that only has 2 possible outcomes.
- Sample space: $S = \{\text{success}, \text{failure}\}$
- Suppose

$$X(\text{success}) = 1 \text{ and } X(\text{failure}) = 0.$$

- p.m.f. for X:

$$f(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}$$

- X has a *Bernoulli distribution* with *parameter* p.
- $E(X) = p$
- $\text{Var}(X) = p(1 - p)$
- $\sigma_X = \sqrt{p(1 - p)}$
- Alternative notation: $q = 1 - p$

Definition

- Consider a sequence of Bernoulli trials such that
  - $n =$ the number of trials
  - $p =$ the probability of success on each trial
  - the trials are independent
  - $X =$ number of success that occur
- X has a *binomial distribution* with *parameters* n and p.
- $X \sim b(n, p)$
- p.m.f. for X:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \ldots, n.$$

- $E(X) = np$
- $\text{Var}(X) = np(1-p)$

Definition

The *cumulative distribution function* of X is

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

Often, it is simply called the *distribution function* of X.

Example

There is a 15% chance that items produced in a certain factory are defective. Assuming that 9 items are produced, and assuming that they are statistically independent, what is the probability that

- at most 4 are defective?
- at least 6 are defective?
- more than 6 are defective?
- the number of defective items is between 2 and 5 inclusive?

Connection Between the Hypergeometric and Binomial Distributions

Random Sampling

- Without replacement: hypergeometric
- With replacement: binomial

Example

In a university organization with 200 members, 60 are seniors. In a random sample of size 10, what is the probability that 4 are seniors, if the sampling is done

- without replacement?
- with replacement?
- Find the expected value, variance, and standard deviation of the number of seniors in the sample under both types of sampling.

Outline

Definition

The *moment-generating function* of X is

$$M(t) = E(e^{tX}),$$

assuming $E(e^{tX})$ is finite on some open interval $-h < t < h$.

Example

Let X be a random variable with p.m.f.

$$f(x) = \frac{1}{14}x^2, \text{ for } x = 1, 2, 3.$$

Find the moment generating function of X.

Example

- If the p.m.f. of X is $f(x) = \frac{6}{\pi^2 x^2}$, for $x = 1, 2, \ldots$
- then X does not have a moment generating function.

Example

- Suppose the m.g.f. of X is $M(t) = \frac{3}{6}e^t + \frac{2}{6}e^{2t} + \frac{1}{6}e^{3t}$.
- Find the p.m.f. of X.

Example

Find the p.m.f. of X if the m.g.f. is

$$M(t) = \frac{e^t/2}{1 - e^t/2}, \ t < \ln(2).$$

Theorem

X and Y have the same m.g.f. if and only if they have the same p.m.f.

- $E(X) = M'(0)$
- $E(X^2) = M''(0)$
- $E(X^r) = M^{(r)}(0)$
- $\mathrm{Var}(X) = M''(0) - M'(0)^2$

Review for Exam Two

- Discrete Random Variables
  - Definitions of random variables, discrete random variables, p.m.f., and support.
  - Probabilities involving random variables
  - Properties of a p.m.f.
  - Hypergeometric distribution
- Mathematical Expectation
  - Definition
  - Calculating $E(X)$
  - Properties of E
  - Expected value hypergeometric random variable: $E(X) = np$

- The Mean, Variance, and Standard Deviation
  - Definitions/notation for mean, variance, and standard deviation
  - for random variables and samples
  - Shortcut formulas for variance of a random variable/sample
  - Be able to compute everything “by hand”.
  - Variance of a hypergeometric random variable: $\text{Var}(X) = np(1-p)\frac{N-n}{N-1}$.
  - Property of variance: $\text{Var}(aX + b) = a^2\text{Var}(X)$, if a and b are constants.
- Bernoulli Trials and the Binomial Distribution
  - Probabilities
  - p.m.f.
  - c.d.f. table/computer/calculator
  - $E(X) = np$, $\text{Var}(X) = np(1-p)$

- Moment-Generating Functions
  - p.m.f. $\rightarrow$ m.g.f.
  - m.g.f. $\rightarrow$ p.m.f.
  - $E(X^r) = M^{(r)}(0)$
  - $\text{Var}(X) = M''(0) - M'(0)^2$

Outline

Approximate Poisson Process with Parameter $\lambda > 0$

Setting

- Measuring occurrences of some event on a continuous interval.
- Examples:
  - Number of phone calls received in 1 hour
  - Number of defects in 1 meter of wire

Assumptions

- Occurrences in non-overlapping intervals are independent.
- In a sufficiently short interval of length h, the probability of 1 occurrence is approximately λh.
- In a sufficiently short interval, the probability of 2 or more occurrences is essentially zero.

Poisson Distribution

- Let $X = \#$ of occurrences in an interval of length 1
- Then X has a *Poisson distribution* with parameter λ.
- $$f(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Example

Phone calls received by a company are a Poisson process with parameter $\lambda = 4$. In a 1 minute period, find the probability of receiving

- 2 calls?
- 5 calls?
- at most 3 calls?
- at least 7 calls?

- If X has a Poisson distribution with parameter λ, then
- $E(X) = \text{Var}(X) = \lambda$
- For a Poisson process with parameter λ,
- λ is the average # of occurrences in an interval of length 1.

Example

Phone calls received by a company are a Poisson process, and the company receives an average of 4 calls per minute. In a 3 minute period, find the probability of the company receiving

- 10 calls?
- at most 15 calls?

Interval of Length t

- Consider a Poisson process with parameter λ
- If $X =$ # of occurrences in an interval of length t,
- then X has a Poisson distribution with mean λt.

Example

On average, there are 3 flaws in 8 meters of copper wire. For a piece of wire 20 meters long, find the probability of observing

- 5 flaws.
- fewer than 9 flaws.
- Find the expected value, variance, and standard deviation of the number of flaws on a 20 meter piece of wire.

Data

Example

Let X equal the number of green m&m's in a package of size 22. Forty-five observations of X yielded the following frequencies for the possible outcomes of X:

Outcome (x):	0	1	2	3	4	5	6	7	8	9
Frequency:	0	2	4	5	7	9	8	5	3	2

- Calculate $\bar{x}$ and s^2. Are they close?
- Compare the relative frequency histogram to the probability histogram of a Poisson random variable with mean $\lambda = 5$.
- Do these data appear to be observations from a Poisson random variable?

Poisson Approximation to the Binomial

- If n is large and p is small,
- $\text{bin}(n, p) \approx$ Poisson with $\lambda = np$.
- $n \geq 100$ and $p \leq 0.1$
- $n \geq 20$ and $p \leq 0.05$

Example

In a shipment of 2000 items, 4% are defective. In a random sample of size 100, find the approximate probability that more than 10 items are defective.