CSCI 3022 Intro to Data Science
nb0708 (Load 'em up!)

Example:

A factory makes parts for a medical device company. 6% of those parts are defective. For each one of the problems below:

(i.) Define an appropriate random variable for the experiment.

(ii.) Give the values that the random variable can take on.

(ii.) Find the probability that the random variable equals 2.

(iv.) State any assumptions you need to make.

Problems:

1. Out of 10 parts, X are defective.
2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.
3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

Announcements and Reminders

- Work on HW 3
- Past Exams and solutions posted to Canvas!
- Exam is untimed, takehome. Designed to take around 2-3 hours, but you'll have most of a week. Due Mar 8, **likely** posted Mar 2. Think of it as an "all pen-and-paper" homework. Submission via Gradescope:
  1. You can TeX it yourself (template will be provided)
  2. You can take the given pdf and annotate on it
  3. You can print the pdf and write on it, then picture/scan your solutions
  4. You can work the exam on regular paper and picture/scan your work
- Tested content through lecture Mar 1 (Expectation & Variance)

Last Time...: the blocks of discrete probability

1. Bernoulli: *one* binary outcome experiment.
2. Binomial: binary outcome experiment success *count* in n tries.

   exactly n flips
3. Geometric: Total trials *until a success* of a binary outcome experiment.

   we don't know "n"; that's what we're estimating
4. Negative Binomial: Trials until r binary outcome experiment *successes*.

   r geometrics in a row.
5. Poisson: *counting* outcomes with a fixed rate λ.

   not based on trial-and-error coin flips

Last Time...: the blocks of discrete probability

$\binom{10}{2} = \frac{10!}{2!(10-2)!} = \frac{10!}{2!8!}$

1. Bernoulli: *one* binary outcome experiment.
   $f(x) = p^x(1-p)^{1-x}$
2. Binomial: binary outcome experiment success *count* in n tries. $\binom{10}{8} = \frac{10!}{8!(10-8)!}$
   $f(x) = \boxed{\binom{n}{x}} p^x(1-p)^{(n-x)}$
3. Geometric: Total trials *until a success* of a binary outcome experiment.
   $f(x) = (1-p)^{x-1}p$
4. Negative Binomial: Trials until *r* binary outcome experiment *successes*.
   $f(x) = \binom{x-1}{r-1}p^r(1-p)^{(x-r)}$
5. Poisson: *counting* outcomes with a fixed rate λ.
   $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Last Time...: the blocks of discrete probability

The underlying pieces of discrete RVs:

1. The random variable X takes inputs/events in the (discrete) sample space Ω and maps them to a (discrete) finite or infinite set of probability values $a_1, a_2, a_3, \ldots$.
2. We find probabilities in the probability mass function or probability density function

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

3. We can find cumulative probabilities or probability on ranges of outcomes in the cumulative density function

$$F(x) = P(X \leq x) = \sum_{X \leq x} f(x).$$

Discrete Distributions Example

Example:
A factory makes parts for a medical device company. 6% of those parts are defective. For each one of the problems below:

(i.) Define an appropriate random variable for the experiment.

(ii.) Give the values that the random variable can take on.

(ii.) Find the probability that the random variable equals 2.

(iv.) State any assumptions you need to make.

Problems:

1. Out of 10 parts, X are defective.
2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.
3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

Discrete Distributions Example

Part 1 ; Part 2 ; Part 3 ; ... ; Part 10
Y/N ; Y/N ; Y/N ; ... ; Y/N

6% of those parts are defective.

1. Out of 10 parts, X are defective.

fixed

(i.) r.v.: $X \sim bin(\underset{n}{10}, \underset{p}{0.06})$

"X is distributed as binomial w/ 10 trials & prob of success of .06".

(ii.) Values of r.v.: $X \in \{0, 1, 2, 3, \ldots, 10\}$

(iii.) $P(X = 2)$:

(iv.) Assumptions:

$$P(X=2) = \underbrace{\binom{10}{2}}_{\text{\# of orders}} \cdot \underbrace{(.06)^2 (1-.06)^8}_{P(\text{one such order})}$$

Discrete Distributions Example

6% of those parts are defective.

1. Out of 10 parts, X are defective.

(i.) r.v.: $X \sim bin(10, .06)$

(ii.) Values of r.v.: $X \in \{0, 1, 2, \ldots, 10\}$

(iii.) $P(X = 2)$: $\binom{10}{2}.06^2.94^8$

(iv.) Assumptions: Parts are *i.i.d.*

independent & identically distributed

Bernoullis.

Discrete Distributions Example

6% of those parts are defective.

2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.

(i.) r.v.:

(ii.) Values of r.v.:

(iii.) $P(X = 2)$:

(iv.) Assumptions:

Discrete Distributions Example

6% of those parts are defective.

2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.

(i.) r.v.: $X + 1 \sim \boxed{Geom(.06)}$ OR negative binomial w/ "r = 1"

(ii.) Values of r.v.: $X \in \{0, 1, 2, \ldots, \infty\}$

(iii.) $P(X = 2)$: $.94^2 .06^1$

(iv.) Assumptions: Parts are *i.i.d.*

Discrete Distributions Example

6% of those parts are defective.

count

3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

(i.) r.v.:

(ii.) Values of r.v.:

(iii.) $P(X = 2)$:

(iv.) Assumptions:

Discrete Distributions Example

6% of those parts are defective.

3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

(i.) r.v.: $X \sim Pois(10)$ silly result.

(ii.) Values of r.v.: $X \in \{0, 1, 2, \ldots, \infty\}$

(iii.) $P(X = 2)$: $\frac{e^{-10} \cdot 10^2}{2!}$

(iv.) Assumptions: Parts are... *Poisson*?