STAT509: Discrete Random Variable

Peijie Hou

University of South Carolina

September 16, 2014

Motivation

- So far, we have already known how to calculate probabilities of events.
- Suppose we toss a fair coin three times, we know that the probability of getting three heads in a row is $1/8$. However, people maybe not interested in the probability directly related to the sample space, instead, some numerical summaries are of our interest.
- Suppose we are interested in the number of heads in a experiment of tossing a fair coin three times. The possible values are 0, 1, 2, and 3. The corresponding probabilities can be calculated.

Random Variable

Definition

Let S be the sample space of a random experiment. A **random variable** defined on S is a function from S into the reals. We usually denote random variables by the capital letters X, Y , etc. In mathematical notation,

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

The set of possible distinct values of the random variable is called its **range**. A realized value of the random variable is usually denoted by the lower-case letters x, y,etc.

Type of Random Variables

- A **discrete** random variable can take one of a countable list of distinct values. Usually, counts are discrete.
- A **continuous** random variable can take any value in an interval of the real number line. Usually, measurements are continuous.
- Classify the following random variables as discrete or continuous
  - Time until a projectile returns to earth.
  - The number of times a transistor in computer memory changes state in one operation.
  - The volume of gasoline that is lost to evaporation during the filling of a gas tank.
  - The outside diameter of a machined shaft.

Example of Discrete Random Variable

- Consider toss a fair coin 10 times. The sample space S contains total $2^{10} = 1024$ elements, which is of the form

$$S = \{TTTTTTTTTT, \ldots, HHHHHHHHHH\}$$

- Define the random variable Y as the number of tails out of 10 trials. Remeber that a random variable is a map from sample space to real number. For instance,

$$Y(TTTTTTTTTT) = 10.$$

- The range (all possible values) of Y is $\{0, 1, 2, 3, \ldots, 10\}$, which is much smaller than S.

Example: Mechanical Components

- An assembly consists of three mechanical components. Suppose that the probabilities that the first, second, and third components meet specifications are 0.90, 0.95 and 0.99 respectively. Assume the components are independent.
- Define event
  S_i = the i^{th} component is within specification, where $i = 1, 2, 3$.
- One can calculate
  $P(S_1 S_2 S_3) = (0.9)(0.95)(0.99) = 0.84645$,
  $P(S_1 \overline{S_2} \overline{S_3}) = (0.9)(0.05)(0.01) = 0.00045$, etc.

Example: Mechanical Components

Possible Outcomes for One assembly is

$S_1S_2S_3$	$S_1\overline{S_2S_3}$	$\overline{S_1}S_2\overline{S_3}$	$\overline{S_1S_2}S_3$	$S_1S_2\overline{S_3}$	$S_1\overline{S_2}S_3$	$\overline{S_1}S_2S_3$	$\overline{S_1S_2S_3}$
0.84645	0.00045	0.00095	0.00495	0.00855	0.04455	0.09405	0.00005

Let Y = Number of components within specification in a randomly chosen assembly. Can you fill in the following table?

Y	0	1	2	3
$P(Y = y)$				

Solution:

Probability Mass Functions of Discrete Variables

- Definition: Let X be a discrete random variable defined on some sample space S. The **probability mass function** (pmf) associated with X is defined to be

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

- A pmf $p(x)$ for a discrete random variable X satisfies the following:
  1. $0 \leq p(x) \leq 1$, for all possible values of x.
  2. The sum of the probabilities, taken over all possible values of X, must equal 1; i.e.,

$$\sum_{\text{all } x} p(x) = 1.$$

Cumulative Distribution Function

- Definition: Let Y be a random variable, the **cumulative distribution function** (CDF) of Y is defined as

$$F_Y(y) \equiv P(Y \leq y).$$

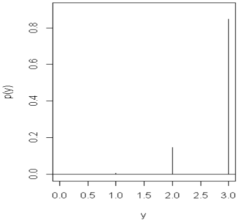
- $F(y) \equiv P(Y \leq y)$ is read, "the probability that the random variable Y is less than or equal to the value y."
- Property of cumulative distribution function
  1. $F(y)$ is a nondecreasing function.
  2. $0 \leq F_Y(y) \leq 1$, since $F_Y(y) = P(Y \leq y)$ is a probability.

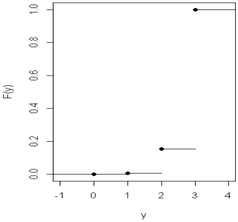
Example: Mechanical Components Revisited

- In the Mechanical Components example Y = Number of components within specification in a randomly chosen assembly. We already filled in the following table:

Y	0	1	2	3
$P(Y = y)$	0.00005	0.00635	0.14715	0.84645

- The pmf and CDF plots of Y





Expected Value of a Random Variable

- Definition: Let Y be a discrete random variable with pmf $p_Y(y)$, the **expected value** or **expectation** of Y is defined as

$$\mu = \mathrm{E}(Y) = \sum_{\text{all } y} y p_Y(y).$$

  The expected value for a discrete random variable Y is simply a weighted average of the possible values of Y. Each value y is weighted by its probability $p_Y(y)$.

- In statistical applications, $\mu = \mathrm{E}(Y)$ is commonly called the **population mean**.

Expected Value: Mechanical Components

- The pmf of Y in the mechanical components example is

Y	0	1	2	3
$P(Y=y)$	0.00005	0.00635	0.14715	0.84645

- The expected value of Y is

$$\begin{aligned}\mu &= \mathrm{E}(Y) = \sum_{\text{all } y} y p_Y(y) \\ &= 0(.00005) + 1(.00635) + 2(.14715) + 3(.84645) \\ &= 2.84\end{aligned}$$

- Interpretation: **On average**, we would expect 2.84 components within specification in a randomly chosen assembly.

Expected Value as the Long-run Average

- The expected value can be interpreted as the long-run average.
- For the mechanical components example, if we randomly choose an assembly, and record the number of components within specification. Over the **long run**, the **average** of these Y observation would be close (converge) to 2.84.

Expected Value of Functions of Y

- Definition: Let Y be a discrete random variable with pmf $p_Y(y)$. Let g be a real-valued function defined on the range of Y. The **expected value** or **expectation** of $g(Y)$ is defined as

$$\mathrm{E}[g(Y)] = \sum_{\text{all } y} g(y) P_Y(y).$$

- Interpretation: The expectation $\mathrm{E}[g(Y)]$ could be viewed as the weighted average of the function g when evaluated at the random variable Y.

Properties of Expectation

Let Y be a discrete random variable with pmf $p_Y(y)$. Suppose that $g_1, g_2, \ldots, g_k$ are real-valued function, and let c be a real constant. The expectation of Y satisfies the following properties:

1. $\mathrm{E}[c] = c$.
2. $\mathrm{E}[cY] = c\mathrm{E}[Y]$.
3. Linearity: $\mathrm{E}\left[\sum_{j=1}^{k} g_j(Y)\right] = \sum_{j=1}^{k} \mathrm{E}[g_j(Y)]$.

Remark: The 2nd and 3rd rule together imply that

$$\mathrm{E}\left[\sum_{j=1}^{k} c_j g_j(Y)\right] = \sum_{j=1}^{k} c_j \mathrm{E}[g_j(Y)],$$

for constant c_j.

Example: Mechanical Components

Question: In the Mechanical Components example,
Y = Number of components within specification in a randomly chosen assembly.
We calculated the expected value of Y $\mathrm{E}(Y) = 2.84$. Suppose that the cost (in dollars) to repair the assembly is given by the cost function $g(Y) = (3 - Y)^2$.
What is the expected cost to repair an assembly?

Solution: In order to use the linearity property, expanding the square term gives $g(Y) = 9 - 6Y + Y^2 = Y^2 - 6Y + 9$. The expected cost

$$\mathrm{E}[g(Y)] = \mathrm{E}[Y^2 - 6Y + 9] = \mathrm{E}[Y^2] - 6\mathrm{E}[Y] + 9.$$

Example: Mechanical Components

The expected value of Y^2 is

$$
\begin{aligned}
\mathrm{E}[Y^2] &= \sum_{\text{all } y} y^2 p_Y(y) \\
&= 0^2(.00005) + 1^2(.00635) + 2^2(.14715) + 3^2(.84645) \\
&= 8.213
\end{aligned}
$$

It follows that the average cost in dollars equals

$$\mathrm{E}[g(Y)] = 8.213 - 6(2.84) + 9 = 0.173.$$

Interpretation: One average, the repair cost on each assembly is \$0.173.

Example: Project Management

A project manager for an engineering firm submitted bids on three projects. The following table summarizes the firms chances of having the three projects accepted.

Project	A	B	C
Prob of accept	0.30	0.80	0.10

Question: Assuming the projects are independent of one another, what is the probability that the firm will have all three projects accepted?

Solution:

Example: Project Management Cont'd

Question: What is the probability of having at least one project accepted?

Project	A	B	C
Prob of accept	0.30	0.80	0.10

Solution:

Example: Project Management Cont'd

Question: Let Y = number of projects accepted. Fill in the following table and calculate the expected number of projects accepted.

Y	0	1	2	3
$p_Y(y)$	0.126			

Solution:

Variance of a Random Variable

- Definition: Let Y be a discrete random variable with pmf $p_Y(y)$ and expected value $\mathrm{E}(Y) = \mu$. The **variance** of Y is given via

$$\sigma^2 \equiv \mathrm{Var}(Y) \equiv \mathrm{E}\left[(Y-\mu)^2\right] = \sum_{\text{all } y}(y-\mu)^2 p_Y(y).$$

Warning: Variance is **always** non-negative!

- Definition: The **standard deviation** of Y is the positive square root of the variance:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\mathrm{Var}(Y)}.$$

Remarks on Variance (σ^2)

Suppose Y is a random variable with mean μ and variance σ^2.

- The variance is the average **squared distance** from the mean μ.
- The variance of a random variable Y is a measure of **dispersion** or **scatter** in the possible values for Y.
- The larger (smaller) σ^2 is, the more (less) spread in the popssible values of Y about the population mean $\mathrm{E}(Y)$.
- **Computing Formula:**
$\mathrm{Var}(Y) = \mathrm{E}[(Y-\mu)^2] = \mathrm{E}[Y^2] - [E(Y)]^2$.

Properties on Variance (σ^2)

Suppose X and Y are random variable with finite variance. Let c be a constant.

- $\mathrm{Var}(c) = 0$.
- $\mathrm{Var}[cY] = c^2\mathrm{Var}[Y]$.
- Suppose X and Y are independent, then

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y].$$

Question: $\mathrm{Var}[X - Y] = \mathrm{Var}[X] - \mathrm{Var}[Y]$.

- TRUE
- FALSE.

Bernoulli Trials

Definition: Many experiments can be envisioned as consisting of a sequence of "trials," these trails are called Bernoulli trails if

1. Each trial results in only two possible outcomes, labelled as "success" and "failure".
2. The trials are independent.
3. The probability of a success in each trial, denoted as p, remains constant.

Bernoulli Trials

- By convention, a success is recoded as "1," and a failure as "0."
- Define X to be the Bernoulli random variable
  - $X = 1$ represents "success"; $X = 0$ represents "failure"
  - Suppose $P(X = 1) = p$ and $P(X = 0) = 1 - p$, then

$$E(X) = p \text{ and } \mathrm{Var}(X) = pq$$

Proof:

Examples of Bernoulli Trials

- Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant.
  - ✓ Detecting water pollutant="trial"
  - ✓ polluted water is observed="success"
  - ✓ $p = P(\text{"success"}) = P(\text{polluted water}) = 0.1$.

- Ninety-eight percent of all air traffic radar signals are correctly interpreted the first time they are transmitted.
  - ✓ radar signal = "trial"
  - ✓ signal is correctly interpreted="success"
  - ✓ $p = P(\text{"success"}) = P(\text{correct interpretation}) = 0.98$.

Binomial Distribution

- Definition: Suppose that n Bernoulli trials are performed. Define

  $$Y = \text{the number of successes (out of } n \text{ trials performed).}$$

  We say that the random variable Y has a **binomial distribution** with number of trials n and success probability p.

- Shorthand notation is $Y \sim b(n, p)$.
- Let us derive the pmf of Y through an example.

Example: Water Filters

- A manufacturer of water filters for refrigerators monitors the process for defective filters. Historically, this process averages 5% defective filters. Five filters are randomly selected.
- Find the probability that no filters are defective.
  solution:
- Find the probability that exactly 1 filter is defective.
  solution:
- Find the probability that exactly 2 filter is defective.
  solution:
- Can you see the pattern?

Probability Mass Function of Binomial R.V.

- Suppose $Y \sim b(n, p)$.
- The pmf of Y is given by

$$p(y) = \begin{cases} \binom{n}{y} p^y (1-p)^{n-y} & \text{for } y = 0, 1, 2, \ldots, n, \\ 0 & \text{otherwise.} \end{cases}$$

- Recall that $\binom{n}{r}$ is the number of ways to choose r distinct **unordered** objects from n distinct objects:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

Mean and Variance of Binomial R.V.

- If $Y \sim b(n, p)$, then

$$\mathrm{E}(Y) = np$$
$$\mathrm{Var}(Y) = np(1-p).$$

- (Optional) I will derive above formulas on the blackboard.
  Proof:

Example: Radon Levels

Question: Historically, 10% of homes in Florida have radon levels higher than that recommended by the EPA. In a random sample of 20 homes, find the probability that exactly 3 have radon levels higher than the EPA recommendation. Assume homes are independent of one another.

Solution: Checking each house is a Bernoulli trial, which satisfies

- ✓ Two outcomes: higher than EPA recommendation or satisfies EPA recommendation.
- ✓ Homes are independent of one another.
- ✓ The case that the radon level is higher than the recommendation is considered as a "success". The success probability remains 10%.

Example: Radon Levels Cont'd

So, the binomial distribution is applicable in this example. Define

$Y =$ number of home having radon level higher than EPA.

We have $Y \sim b(20, 0.1)$.

$$P(Y = 3) = \binom{20}{3} 0.1^3 0.9^{20-3} = 0.1901.$$

Doing calculation by `R` is much easier:

```
> dbinom(3,20,0.1)
[1] 0.1901199
```

Example: Radon Levels Cont'd

You can also calculate the probability that at least 6 homes out of the sample having higher radon level than recommended:

$$P[Y \geq 6] = 1 - P[Y \leq 5].$$

Using R,

```
> 1-pbinom(5,20,0.1)
[1] 0.01125313
```

It's Your Turn...

If a manufacturing process has a 0.03 defective rate, what is the probability that at least one of the next 25 units inspected will be defective?

(a) $\binom{25}{1}(0.03)^1(0.97)^{24}$

(b) 1-$\binom{25}{1}(0.03)^1(0.97)^{24}$

(c) 1-$\binom{25}{0}(0.03)^0(0.97)^{25}$

(d) 1-$\binom{25}{0}(0.03)^{25}(0.97)^0$

Question

A manufacturing process has a 0.03 defective rate. If we randomly sample 25 units

(a) What is the probability that less than 6 will be defective?

(b) What is the probability that 4 or more are defective?

(c) What is the probability that between 2 and 5, inclusive, are defective?

Geometric Distribution

Setting: Suppose that Bernoulli trials are continually observed. Define

$$Y = \text{the number of trials to observe the } \textbf{first} \text{ success.}$$

Then we say that Y has a geometric distribution with success probability p, where p is the success probability of Bernoulli trials.

- Shorthand notation is $Y \sim \text{geom}(p)$.
- The pmf of Y is

$$P(Y = y) = p(1-p)^{y-1} \quad x = 1, 2, \ldots$$

Example: Geometric Distribution

The probability that a wafer contains a large particle of contamination is 0.01. If it is assumed that the wafers are independent, what is the probability that exactly 125 wafers need to be analysed until a large particle is detected?

Solution: Let X denote the number of samples analysed until a large particle is detected. Then X is a geometric random variable with $p = 0.01$. The requested probability is

$$P(X = 125) = (0.01)(0.99)^{125-1} = 0.0029$$

Mean and Variance of Geometric Distribution

Suppose $Y \sim \text{geom}(p)$, the expected value of Y is

$$\mathrm{E}(Y) = \frac{1}{p}.$$

Proof.

Let $q = 1 - p$, $E(Y) = \sum_{y=1}^{\infty} yq^{y-1}p = p\sum_{y=1}^{\infty} yq^{y-1}$. The right-hand side is the derivative of $p\sum_{y=1}^{\infty} q^y$ w.r.t. q. By geometric series,

$$p\sum_{y=1}^{\infty} q^y = \frac{pq}{1-q}.$$

It follows that $\mathrm{E}(Y) = \frac{\partial}{\partial q}\left[\frac{pq}{1-q}\right] = \frac{1}{p}$ □

Mean and Variance of Geometric Distribution

The Variance can be derived in a similar way, which is given by

$$\mathrm{Var}\,(Y) = \frac{1-p}{p^2}$$

Golf Resort

The probability that a wafer contains a large particle of contamination is 0.01. If it is assumed that the wafers are independent, then on average, how many wafers need to be analysed until a large particle is detected?

Solution:

Hypergeometric Distribution

Setting: A set of N objects contains r objects classified as successes $N - r$ objects classified as failures. A sample of size n objects is selected randomly (**without replacement**) from the N objects. Define

$$Y = \text{the number of success (out of the } n \text{ selected).}$$

We say that Y has a **hypergeometric distribution** and write $Y \sim \text{hyper}(N, n, r)$. The pmf of Y is given by

$$p(y) = \begin{cases} \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}, & y \leq r \text{ and } n - y \leq N - r \\ 0, & \text{otherwise.} \end{cases}$$

Mean and Variance of Hypergeometric Distribution

If $Y \sim \text{hyper}(N, n, r)$, then

- $E(Y) = n(\frac{r}{N})$
- $\mathrm{Var}(Y) = n(\frac{r}{N})(\frac{N-r}{N})(\frac{N-n}{N-1})$.

Example: Hypergeometric Distribution

A batch of parts contains 100 parts from a local supplier of tubing and 200 parts from a supplier of tubing in the next state. If four parts are selected randomly and without replacement, what is the probability they are all from the local supplier?

Solution: Let us identify the parameters first. In this example, $N =$, $r =$, and $n =$.
Let X equal the number of parts in the sample from the local supplier. Then,

$$P(X = 4) = \frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}} = 0.0119$$

Example: Hypergeometric Distribution

The R code to calculate $P(Y = y)$ is of the form dhyper(y,r,N-r,n). Using R,

```
> dhyper(4,100,200,4)
[1] 0.01185408
```

What is the probability that two or more parts in the sample are from the local supplier?

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.5925943,$$

where $P(X \leq 1)$ can be computed via R:

```
> phyper(1,100,200,4)
[1] 0.5925943
```

It's your turn...

If a shipment of 100 generators contains 5 faulty generators, what is the probability that we can select 10 generators from the shipment and not get a faulty one?

(a) $\frac{\binom{5}{1}\binom{95}{9}}{\binom{100}{10}}$

(b) $\frac{\binom{5}{0}\binom{95}{9}}{\binom{100}{9}}$

(c) $\frac{\binom{5}{0}\binom{95}{10}}{\binom{100}{10}}$

(d) $\frac{\binom{5}{10}\binom{95}{0}}{\binom{100}{10}}$

Insulated Wire

- Consider a process that produces insulated copper wire. Historically the process has averaged 2.6 breaks in the insulation per 1000 meters of wire. We want to find the probability that 1000 meters of wire will have 1 or fewer breaks in insulation?
- Is this a binomial problem?
- Is this a hypergeometric problem?

Poisson Distribution

Note: The Poisson distribution is commonly used to model counts, such as

1. the number of customers entering a post office in a given hour
2. the number of α-particles discharged from a radioactive substance in one second
3. the number of machine breakdowns per month
4. the number of insurance claims received per day
5. the number of defects on a piece of raw material.

Poisson Distribution Cont'd

- Poisson distribution can be used to model the number of events occurring in a continuous time or space.
- Let λ be the **average number** of occurrences per **base unit** and t is the number of base units inspected.
- Let Y = the number of "occurrences" over in a unit interval of time (or space). Suppose Poisson distribution is adequate to describe Y. Then, the pmf of Y is given by

$$p_Y(y) = \begin{cases} \frac{(\lambda t)^y e^{-\lambda t}}{y!}, & y = 0, 1, 2, \ldots \\ 0, & \text{otherwise.} \end{cases}$$

- The shorthand notation is $X \sim \text{Poisson}(\lambda t)$.

Mean and Variance of Poisson Distribution

If $Y \sim \text{Poisson}(\lambda t)$,

$$\mathrm{E}(Y) = \lambda t$$
$$\mathrm{Var}(Y) = \lambda t.$$

Go Back to Insulated Wire

- "Historically the process has averaged 2.6 breaks in the insulation per 1000 meters of wire" implies that the average number of occurrences $\lambda = 2.6$, and the base units is 1000 meters.
- Let $X=$ number of breaks 1000 meters of wire will have. $X \sim \text{Poisson}(2.6)$.
- We have

$$P(Y = 0 \cup Y = 1) = \frac{2.6^0 e^{-2.6}}{0!} + \frac{2.6^1 e^{-2.6}}{1!}.$$

- Using R,

```
> dpois(0,2.6)+dpois(1,2.6)
[1] 0.2673849
```

Go Back to Insulated Wire Cont'd

- If we were inspecting 2000 meters of wire, $\lambda t = (2.6)(2) = 5.2$.
$$Y \sim \text{Poisson}(5.2).$$
- If we were inspecting 500 meters of wire, $\lambda t = (2.6)(0.5) = 1.3$.
$$Y \sim \text{Poisson}(1.3)$$

.

Conditions for a Poisson Distribution

- Areas of inspection are independent of one another.
- The probability of the event occurring at any particular point in space or time is negligible.
- The mean remains constant over all areas of inspection.

The pmf of Poisson distribution is derived based on these three assumption!
`http://www.pp.rhul.ac.uk/~cowan/stat/notes/PoissonNote.pdf`

Questions

1. Suppose we average 5 radioactive particles passing a counter in 1 millisecond. What is the probability that exactly 3 particles will pass in one millisecond?
   Solution:

2. Suppose we average 5 radioactive particles passing a counter in 1 millisecond. What is the probability that exactly 10 particles will pass in three milliseconds?
   Solution: