STA 111 (Summer Session I)

Lecture Five – Expectation Cont'd; The Normal Distribution Cont'd
D.S. Sections 4.1, 4.2, 4.3, 5.6 and 5.9

Instructor: Olanrewaju (Michael) Akande

Department of Statistical Science, Duke University

May 25, 2016

Outline

- Questions from Last Lecture.
- Expectation Cont'd
- The Normal Distribution Cont'd
- The Normal Approximation to the Binomial Distribution
- The Multinomial Distribution
- Recap

Introduction

– In the last lecture we learned how to calculate expectation and variance.

– We also learned about the standard normal distribution.

– Today we will go over more examples to help understand expectations.

– We will also continue with our discussions on the standard normal distribution and extend the discussion to the arbitrary case. We will move on to look at the normal approximation to the binomial distribution.

– Lastly, we will learn about the generalization of the binomial distribution

Additional Review

First let's a few more things we previously skipped.

1. Recall that two events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. A and B are said to be "conditionally independent" given a third event C if

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C)$$

2. Let $f(x)$ be the pdf, and $F(x)$ be the cdf of a continuous random variable X. Then,

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$$

3. For any random variable, the variance $\mathbb{V}[X] \geq 0$
4. If X and Y are independent random variables, then

$$\mathbb{V}[aX + bY] = a^2\mathbb{V}[X] + b^2\mathbb{V}[Y]$$

Turns out we don't need independence here but let's revisit this later!

Additional Concepts Cont'd

The following are true for any continuous random variable X and constants a and b:

1. $\mathbb{P}(X \leq b) = \mathbb{P}(X < b)$ and $\mathbb{P}(X \geq a) = \mathbb{P}(X > a)$. This is true because we assign zero probability to events such as $X = b$ for continuous random variables, that is $\mathbb{P}(X = b) = 0$
2. $\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a < X < b)$ for the same reason as above.
3. Any of the probabilities in (3) above $= F(b) - F(a)$ where $F(x)$ is the cdf of X.

Expectation of a Discrete Random Variable Cont'd

Example 1 (To be done in class): Consider a random variable X which can take on the values: 1, 2, 3 and 4 and define its probability mass function to be $f(x) = c/x^2$. Also consider another random variable Y which can take on the following values: 5, 10, 15, 20 and 25 and define its probability mass function to be $f(y) = ky$.

Find: (i) c (ii) k (iii) $\mathbb{E}(X)$ (iv) $\mathbb{E}(Y)$ (v) $\mathbb{V}(X)$ (vi) $\mathbb{E}(X - Y)$ (vii) $\mathbb{E}(2X^2 + 3Y + 5)$

Expectation of a Continuous Random Variable

Example 2 (D.S. 4.1.6): An appliance has a maximum lifetime of one year. The time X until it fails is a random variable whose p.d.f is:

$$\begin{cases} 2x & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = 5X^4$. Then,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x(2x)\mathrm{d}x = \int_0^1 2x^2 \mathrm{d}x = \frac{2x^3}{3}\Big|_0^1 = \frac{2}{3}$$

$$\mathbb{E}[Y] = \mathbb{E}[5X^4] = \int_{-\infty}^{\infty} 5x^4(2x)\mathrm{d}x = \int_0^1 10x^5 \mathrm{d}x = \frac{10x^6}{6}\Big|_0^1 = \frac{5}{3}$$

How did we know it was continuous?

Expectation of a Continuous Random Variable Cont'd

Example 3: Suppose that a random variable X has pdf $f(x) = c$ for some constant c, where $1 \leq x \leq 3$. Can we find its expected value and variance such that it doesn't involve c? Of course!

Since the pdf must integrate to 1, we know how to find c. That is,

$$1 = \int_1^3 c \, \mathrm{d}x = cx\Big|_1^3 = 3c - c = 2c \Rightarrow c = \frac{1}{2}$$

Then,

$$\mathbb{E}[X] = \int_1^3 \frac{x}{2} \, \mathrm{d}x = \frac{x^2}{4}\Big|_1^3 = \frac{9}{4} - \frac{1}{4} = 2$$

$$\mathbb{E}[X^2] = \int_1^3 \frac{x^2}{2} \, \mathrm{d}x = \frac{x^3}{6}\Big|_1^3 = \frac{27}{6} - \frac{1}{6} = \frac{26}{6} = 4.333$$

$$\Rightarrow \mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 4.333 - 2^2 = 0.333$$

It turns out that this is another well known distribution. A random variable is said to have a uniform distribution (continous) over its support $a \leq x \leq b$ if $f(x) = c$ for some constant c. This is denoted $X \sim Un(a, b)$.

Properties of Expectation

Now lets review two more interesting properties of expectations.

1. If $X_1, X_2, \ldots, X_n$ are n random variables such that each expectation is finite and well-defined, then

$$\mathbb{E}[X_1 + X_2 + \ldots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_n]$$

2. If $X_1, X_2, \ldots, X_n$ are n "independent" random variables such that each expectation is finite and well-defined, then

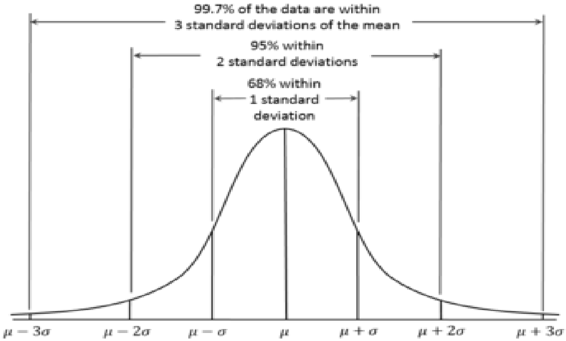
$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}[X_i]$$

Note that we haven't defined what independence means for random variables. We will get to that soon. This is just something to have in mind before then!

The Normal Distribution

– In the last lecture, we started to learn about the normal distribution. This distribution has many nice properties, some of which we already discussed.

– One consequence of those properties is what is called "the empirical rule". The summary of the rule can be seen in the wikipedia image below.



The Normal Distribution Cont'd

– We also learned how to find areas under the standard normal distribution (that is, with mean $\mu = 0$ and variance $\sigma^2 = 1$) using the standard normal cumulative table.

– A region under a normal curve corresponds to a proportion of the population. This is because a normal curve can be viewed as the limit of a series of histograms, in which the sample gets large while the bin-size goes to zero.

– Thus if a student's arrival time in minutes for class is represented by a standard normal, then half the time the student arrives before class starts, and approximately 68% of the time the student is within ± 1 minute of the start of class.

The Normal Distribution Cont'd

– We now show how to convert a question about an arbitrary normal distribution into an equivalent question about the standard normal, and vice-versa. Thus we can use the table to answer questions about all normal distributions, not just the standard normal.

– Let X be a random variable from a normal population with mean μ and variance σ^2. We write this as $X \sim N(\mu, \sigma^2)$. Some textbooks write $X \sim N(\mu, \sigma)$ using the standard deviation instead.

– Define a new random variable $Z = \dfrac{X - \mu}{\sigma}$. Then it turns out that $Z \sim N(0, 1)$. This transformation from X to Z is called the z-transformation. Well this is great! To find probabilities under any normal distribution, we simply have to do the z-transformation to use the standard normal table and we love that don't we?

– To go the other way, we convert the standard normal value to an arbitrary normal distribution by solving for X. So that $X = \mu + Z\sigma$.

The Normal Distribution Cont'd

Example 4: Reggie Jackson, the famous baseball player, has an IQ of 140. What percentage of people are smarter?
Assume that IQs are normally distributed with mean 100 and standard deviation 16.

We want to find $\mathbb{P}(X > 140)$ where $X \sim N(100, 16^2)$. That is, we want the area under the normal distribution for IQ that lies to the right of 140. By the z-transformation, this is equivalent to the area under the standard normal distribution that lies to the right of

$$z = \frac{X - \mu}{\sigma} = \frac{140 - 100}{16} = 2.5.$$

From the normal table, the area above 2.5 is 0.006. Thus about 0.6% of people are smarter than Reggie Jackson.

The Normal Distribution Cont'd

– Now we go the other way. We find the X value that corresponds to a given percentage.

– *Example 5:* To join Mensa one must be in the top 2% of the IQ distribution. What score do you need?

– In the body of the normal table, look up 2%, or 0.02. That gives the z-value of approximately 2.05.

Now we use the inverse z-transformation:

$$X = \mu + Z\sigma = 100 + (2.05)(16) = 132.8.$$

One needs an IQ score of at least 132.8 (i.e., 133) to join.

The Normal Distribution Cont'd

– *Example 6 (To be done in class – D.S. Chapter 5 Exercises, Question 2):* Suppose that X has the normal distribution for which the mean is 1 and variance is 4. Find the value of each of the probabilities:

(a) $\mathbb{P}(X \leq 3)$ (b) $\mathbb{P}(X > 5)$ (c) $\mathbb{P}(X = 1)$ (e) $\mathbb{P}(2 \leq X \leq 5)$

(e) $\mathbb{P}(-1 < X < 0.5)$ (f) $\mathbb{P}(|X| \leq 2)$ (g) $\mathbb{P}(-1 \leq -2X + 3 \leq 8)$

The Normal Approximation to the Binomial

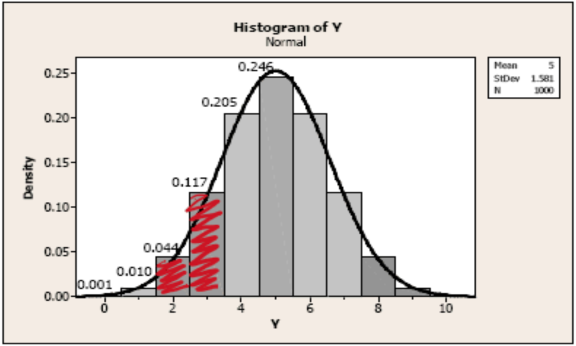
– A perfect normal distribution describes data that can take any possible value – negatives, fractions, irrationals, etc. But often data can only take non-negative integer values.

– In a class of ten students, each tosses a fair coin to decide whether to attend class. So class attendance is a random variable that has the Bin(10, 0.5) distribution. Its mean is $np = 5$ and the standard standard deviation is $\sqrt{n * p * (1 - p)} = 1.581$.

– We can use the normal distribution to estimate the approximate probability that, say, 3 or fewer students will attend tomorrow's lecture. But because only integers are possible, we can improve the accuracy of the normal approximation by using the **continuity correction**.

The Normal Approximation to the Binomial Cont'd

– We approximate the binomial by a normal distribution with the same mean and standard deviation.



– The bad approximation uses the z-transformation $z = (3 - 5)/1.581 = -1.265$, and finds the area under the N(0,1) curve that lies below -1.265 as 0.1020.

The Normal Approximation to the Binomial Cont'd

– The good way handles the area between 3 and 4 appropriately, to take account of the fact that the histogram bar is centered at 3 and we want to include the area up to 3.5 We use the z-transformation $z = (3.5 - 5)/1.581 = -0.949$, and find the probability as 0.1711.

– The normal approximation to the binomial is helpful when n is very large. For example, suppose we wanted to find the probability that more than 20,000 of the 228,330 residents of Durham are unemployed, when the unemployment rate in NC is 10.1%. To use the binomial, we would have to calculate

$$\sum_{x=0}^{20,000} \binom{228,330}{x} (0.101)^x (1 - 0.101)^{228,300-x}.$$

This is intractable, but the normal approximation is not.

– The normal approximation is accurate when $np > 10$ and $n(1 - p) > 10$.

The Multinomial Distribution

– Remember that the Binomial distribution describes a random variable that represents success (or failure) based on n independent trials of an experiments with two outcomes, where the probability of success is the same for each trial. Our toy examples have been tossing a coin n times and rolling a die n times.

– An extension to the binomial distribution which allows for three or more outcomes (k outcomes) for each trial, where the probability of each outcome is the same for each trial is called the Multinomial Distribution.

– A random variable (vector) $\mathbf{X} = (X_1, \ldots, X_k)$ has the multinomial distribution with parameters n and $\boldsymbol{p} = (p_1, \ldots, p_k)$ if its pmf is given by

$$f(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) = \binom{n}{x_1, \ldots, x_k} p_i^{x_1}, \cdots, p_n^{x_k} \quad \text{for} \quad x_1 + \ldots + x_k = n$$

where $\displaystyle \binom{n}{x_1, \ldots, x_k} = \frac{n!}{x_1! x_2! \ldots x_k!}$ is called the multinomial coefficient.

The Multinomial Distribution Cont'd

Example 7 (To be done in class): Suppose that for a single roll of a loaded die we can observe one with probability 0.25, two or three with probability 0.1 each, four or five with probability 0.2 each and six with probability 0.15. What is the probability that in 15 independent rolls, we will observed one four times, two thrice, three once, four once, five four times and six twice.

Let $\boldsymbol{X} = (X_1, X_2, X_3, X_4, X_5, X_6)$ be the observed combination for the die. Then $\boldsymbol{p} = (0.25, 0.1, 0.1, 0.2, 0.2, 0.15)$ and

$$\begin{aligned}
\mathbb{P}\left[\boldsymbol{X} = (4,3,1,1,4,2)\right] &= \binom{n}{x_1, \ldots, x_k} p_i^{x_1}, \cdots, p_n^{x_k} \\
&= \binom{15}{4,3,1,1,4,2} 0.25^4 0.1^3 0.1^1 0.2^1 0.2^4 0.15^2 \\
&= \frac{15!}{4!3!1!1!4!2!} 0.25^4 0.1^3 0.1^1 0.2^1 0.2^4 0.15^2 \\
&= 0.0005
\end{aligned}$$

The Multinomial Distribution Cont'd

A few important points:

1. The vector of probabilities $\boldsymbol{p}$ should sum to 1.
2. Any $X_i \in (X_1, \ldots, X_k)$ has a binomial distribution with parameters n and p_i. *This is why treating the die roll as a binomial when we only care about one of the sides works!*
3. Collapsing the k different outcomes to two outcomes gets you back to a binomial distribution as well. If $\mathbf{X} = (X_1, \ldots, X_k)$ has a multinomial distribution with parameters n and $\boldsymbol{p} = (p_1, \ldots, p_k)$ and $l < k$ where $i_1, \ldots, i_l$ are distinct elements of the set $\{1, \ldots, k\}$, then $Y = X_{i_1} + \cdots + X_{i_l}$ has a binomial distribution with parameters n and $p_{i_1} + \cdots + p_{i_l}$

Recap

We discussed the following:

- Calculating expectations.
- The normal distribution and using it as an approximation for the binomial distribution.
- The multinomial distribution.