STAT 430/510: Lecture 11

James Piette

June 10, 2010
Updates

- HW2 is being graded; should be done over the weekend.
- Pick up your HW1’s up in stat dept. There is a box located right when you enter that is labeled "Stat 430 HW1". It’ll be out for the next week or so.
- Practice midterm will be up sometime tomorrow.
- Next lecture will likely be a combination of some new material, then going into a midterm review. Be sure to come with questions! Otherwise, it’ll be a wasted opportunity.
Uniform R.V.

- **Def:** A continuous r.v. $X$ is said to be uniformly distributed on the interval $[a, b]$ if the pdf of $X$ is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

- The cdf of a uniform r.v. over the interval $[a, b]$ is

$$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

- The important properties of a uniform r.v. are
  - $E(X) = \frac{a+b}{2}$.
  - $Var(X) = \frac{(b-a)^2}{12}$. 
Formalization

- **Def:** $X$ is a **normal** random variable, or $X$ is **normally** distributed, with parameters $\mu$ and $\sigma^2$ if the density of $X$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right) \quad -\infty < x < \infty$$

- We denote this by saying $X \sim N(\mu, \sigma^2)$.

- Note that integrating the density is **not** easy as you need Taylor series (or other tricks) to compute the following (also known as the error function):

$$\text{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2} dt$$
A normal r.v. with $\mu = 0$ and $\sigma^2 = 1$ is known as a standard normal r.v. Below is a plot of its density:
Often, we rewrite the density function of a **standard normal** as the following:

\[ \phi(z) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{z^2}{2} \right) \quad -\infty < z < \infty \]

The cdf of a **standard normal** r.v. is given by:

\[ \Phi(z) = \int_{-\infty}^{z} \phi(s)ds = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds \]

If \( X \sim N(\mu, \sigma^2) \) and \( Z = \frac{X-\mu}{\sigma} \), then \( Z \sim N(0, 1) \).

When doing problems, it is much easier to translate the **normal** r.v. into a **standard normal** and refer to the corresponding table (see Table 5.1 in book).
Properties

- The density of a **normal** r.v. is a bell-shaped curve (i.e. symmetric) around $\mu$, so its density has the following property:

  \[ f(x) = f(\mu + x) \text{ if } x < \mu, \quad f(x) = f(2\mu - x) \text{ if } x > \mu \]

  The same is true for its cdf.

- $E(X) = \mu$.

- $Var(X) = \sigma^2$. 
Example 1

- Here is a basic example.
- Suppose an expert witness in a paternity suit testifies that the length (in days) of human gestation is approximately normally distributed with parameters $\mu = 270$ and $\sigma^2 = 100$. The defendant in the suit is able to prove that he was out of the country during a period that began 290 days before the child’s birth and ended 240 days before the birth.
- **Question**: If the defendant was, in fact, the father, what is the probability that the mother could have a very long or very short gestation as indicated by the testimony?
Example 1 (cont.)

Solution: Let $X$ denote the length of gestation. Then $X$ is normal with parameters $(270, 100)$. So, the probability that the birth could occur within the period is:

$$P(X > 290 \text{ or } X < 240) = P(X > 290) + P(X < 240)$$

$$= P\left(\frac{X - 270}{10} > \frac{290 - 270}{10}\right)$$

$$+ P\left(\frac{X - 270}{10} < \frac{240 - 270}{10}\right)$$

$$= P(Z > 2) + P(Z < -3)$$

$$= (1 - \phi(2)) + \phi(-3)$$

$$= (1 - \phi(2)) + (1 - \phi(3))$$

$$\approx (1 - 0.9772) + (1 - 0.9987) = 0.0241$$
Example 1 (cont.)

To recap what was used to get the final answer in this example:

- First step used the **Axiom** on disjoint events.
- The **normal** r.v. was scaled and shifted in the second step.
- The third step comes directly from the definition of a **standard normal**.
- The fourth step is noticing that $P(Z > 2)$ is the complement of the **standard normal** cdf at 2.
- The final step with $\phi(-3) = (1 - \phi(3))$ can be done since the density of a **normal** is symmetric.
Example 2

- More difficult example.
- Suppose that the travel time from your home to your office is **normally** distributed with mean 40 minutes and standard deviation 7 minutes.

**Question**: If you want to be 95% certain that you will not be late for an office appointment at 1 p.m., what is the latest time you should leave home?

**Solution**: Let $X$ be the travel time. Another way of framing this question is to ask how much time should I give myself so that I arrive on-time with 95% certainty.

So, what is *this* question asking in relation to our $X$?

It is asking what $x$ satisfies the following probability statement:

$$P(X > x) = 0.05$$
Example 2 (cont.)

- That probability statement is equivalent to:
  \[ P \left( \frac{X - 40}{7} > \frac{x - 40}{7} \right) = 0.05 \]
  \[ \Rightarrow P \left( Z > \frac{x - 40}{7} \right) = 0.05 \]

- Looking into the Z-score table, we find out that
  \[ P(Z > 1.645) = 0.05 \]

- So, combining that and our equation above, we get:
  \[ \frac{x - 40}{7} = 1.645 \Rightarrow x = 51.52 \]

- Thus, you should leave no later than 51.52 minutes before 1 p.m. (i.e. 12:08:29).
Example 3

One last example...

The annual rainfall in Cleveland is approximately normally distributed with mean 40.2 inches and standard deviation 8.4 inches. Assume that if \( A_i \) is the event that the rainfall exceeds 44 inches in year \( i \), then the events \( A_i \) are independent.

**Question:** What is the probability that the yearly rainfalls in exactly 3 of the next 7 years will exceed 44 inches?

**Solution:** Very similar to the families and children example. What, then, is the first step?

Find the probability of \( A_i \).
Example 3 (cont.)

Let’s calculate $P(A_i)$:

$$P(A_i > 44) = P\left(\frac{A_i - 40.2}{8.4} < \frac{44 - 40.2}{8.4}\right)$$

$$= P(Z > \frac{3.8}{8.4}) \approx P(Z > 0.45)$$

$$\approx 0.3255$$

Now, we need to determine the probability that exactly 3 out of 7 years will experience at least that much rainfall. What do I use to calculate that?

A binomial r.v., call it $X$, with $n = 7$ and $p = 0.3255$.

$$P(X = 3) = \binom{7}{3}(0.3255)^3(0.6745)^4 = 0.2498$$
**DeMoivre-Laplace Limit Theorem**

**Theorem:** If $S_n$ denotes the number of successes that occur when $n$ independent trials, each with equal probability of success $p$, are performed, then, for any $a < b$,

$$P \left( a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \right) \to \Phi(b) - \Phi(a)$$

as $n \to \infty$. 
Interpretation

Suppose we have a binomial r.v. $X$ with parameters $(n, p)$ (i.e. $n$ trials, probability of success $p$). If we shift our $X$ by $np$ and scale it by $\frac{1}{\sqrt{np(1-p)}}$, then the probability that the altered $X$ lies in an interval between $[a, b]$ can be approximated by taking the difference between the normal cdf of $b$ and the normal cdf of $a$, as $n \to \infty$.

This approximation gets better as $n$ gets bigger; that is, when $n$ is big, the approximation is accurate.

The shifting and scaling is synonymous with what we’ve done in the example above, since $E(X) = np$ and $SD(X) = \sqrt{np(1-p)}$.

Note that $a$ and $b$ do not represent possible outcomes relating to the number of successes. To do that, we must shift and scale the possible successes in the exact same way that we did with $X$. 
Continuity Correction

Before continuing on with examples, we need to discuss a technique needed to better help us approximate.

**Def:** Let $X \sim Bin(n, p)$. If $np(1 - p) > 10$ (our "requirement" to use the approx.), then probabilities of $X$ can be approximated by:

$$P(X \leq a) \approx \Phi \left( \frac{a + 0.5 - np}{\sqrt{np(1 - p)}} \right)$$

$$P(a \leq X \leq b) \approx \Phi \left( \frac{b + 0.5 - np}{\sqrt{np(1 - p)}} \right) - \Phi \left( \frac{a - 0.5 - np}{\sqrt{np(1 - p)}} \right)$$

The 0.5 added/subtracted in each normal cdf represents a correction for the discreteness of a binomial r.v., which is called a **continuity correct.**
Continuity Correction (cont.)

We’ve described how the **continuity correction** works for $P(X \leq a)$ and $P(a \leq X \leq b)$. From those two, we can deduce how the adjustment works for other probabilities like . . .

- $P(X < a)$?
  
  \[
  \Phi \left( \frac{a - 0.5 - np}{\sqrt{np(1-p)}} \right).
  \]

- $P(X > a)$?
  
  \[
  1 - \Phi \left( \frac{a + 0.5 - np}{\sqrt{np(1-p)}} \right).
  \]

- $P(X = a)$?
  
  \[
  \Phi \left( \frac{a + 0.5 - np}{\sqrt{np(1-p)}} \right) - \Phi \left( \frac{a - 0.5 - np}{\sqrt{np(1-p)}} \right).
  \]
Example 4

Let $X$ be the number of times that a fair coin that is flipped 40 times lands on heads.

**Question:** What is the probability that $X = 20$?

**Solution:** Let’s do the normal approximation and after, compare it to the actual.
Example 4 (cont.)

- First, the **normal** approximation:

\[
P(X = 20) = P(19.5 \leq X < 20.5) \text{ (continuity correction)}
\]

\[
= P \left( \frac{19.5 - 20}{\sqrt{10}} < \frac{X - 20}{\sqrt{10}} < \frac{20.5 - 20}{\sqrt{10}} \right)
\]

\[
\approx P(-0.16 < Z < 0.16)
\]

\[
\approx \Phi(0.16) - \Phi(-0.16) \approx 0.1272
\]

- The exact result is:

\[
P(X = 20) = \binom{40}{20} \left( \frac{1}{2} \right)^{40} \approx 0.1254
\]
Example 5

To determine the effectiveness of a certain diet in reducing the amount of cholesterol in the bloodstream, 100 people are put on the diet. After they have been on the diet for a significant length of time, their cholesterol count will be taken. The nutritionist running the experiment decided to endorse the diet if at least 65% of people have a lower cholesterol count after going on the diet.

**Question**: What is the probability that the nutritionist endorses the new diet if, in fact, it has no effect on cholesterol level?
Example 5 (cont.)

- **Solution**: Since it is assumed that the diet has no effect on a person’s cholesterol level, then we can assume the probability that a person’s level would decrease to be . . . 0.50.

- Let $X$ be the number of people whose count is lowered.

$$
P(X \geq 65) = \sum_{i=65}^{100} \binom{100}{i} \left(\frac{1}{2}\right)^{100}$$

$$= P(X \geq 64.5)$$

$$= P\left(\frac{X - (100)(1/2)}{\sqrt{100(1/2)(1/2)}} \geq 2.9\right)$$

$$= 1 - \Phi(2.9) \approx 0.0019$$
Formalization

- **Def**: A continuous r.v. $X$ is said to be **exponentially** distributed with parameter $\lambda$ if the pdf of $X$ is

$$f(x) = \begin{cases} 
\lambda e^{-\lambda x} & \text{if } x \geq 0 \\
0 & \text{if } x < 0
\end{cases}$$

- The cdf for that same **exponential** r.v. is

$$F(x) = \begin{cases} 
1 - e^{-\lambda x} & \text{if } x \geq 0 \\
0 & \text{if } x < 0
\end{cases}$$
Properties

- The moments relating to an **exponential** r.v. are
  - \( E(X) = \frac{1}{\lambda} \).
  - \( Var(X) = \frac{1}{\lambda^2} \).

- **Def:** **Exponential** r.v.’s uphold a unique property known as **memorylessness**, which is that for a r.v. \( X \),

\[
P(X > s + t \mid X > t) = P(X > s) \quad s, t \geq 0
\]

That is, the probability that your r.v. gets to at least \( s + t \), given that it has already reached \( s \), is just the probability that it gets to at least \( s \).
Example 6

- Suppose that the number of miles that a car can run before its battery wears out is **exponentially** distributed with an average value of 10,000 miles.

- **Question**: If a person wants to take a 5000-mile trip, what is the probability that they will be able to complete the trip without having to replace the car battery?

- **Solution**: Since the remaining lifetime is **exponentially** distributed with parameter $\frac{1}{10}$ (taken from the fact that $E(X) = 10000 = \frac{1}{100}$), we can use the memoryless property to say that:

$$P(lifetime > 5000) = 1 - F(5) = e^{-5\lambda} \approx 0.604$$
Example 6 (cont.)

If the distribution of the probability that a battery wears out is not exponential, then things would get a little more complicated:

\[ P(lifetime > t + 5 | lifetime > t) = \frac{1 - F(t + 5)}{1 - F(t)} \]

where \( t \) is the number of miles that the battery had been in use prior to the start of the trip.
(If time permits), let’s look at self-test problem 3.14.

We’ve now covered up to section 5.4.

The exam will cover material up to Ch 4.

As mentioned yesterday, HW3 will be up soon, as well as practice midterm and HW2 solutions.