Math 3215: Lecture 12

Will Perkins

February 23, 2012

1 Review of Statistics

One of you asked a good question about calculating p-values when we introduced the statistical method: 'Does *at least as extreme* mean all outcomes which have probability less than or equal to the outcome we observed?'

That was a good question and made me realize that I hadn't given a precise definition of how to compute a p-value (and actually most textbooks don't have a precise definition of what 'more extreme' means either). So here's a precise definition and procedure for caclulating p-values:

- After specifying H_0 and H_1, decide upon a test statistic T, a random variable that depends on your probability model. In testing a coin for bias, the natural choice for T is the number of heads. In testing whether the average height of a student is at least 6 ft, the average height of a sample of students would be the right choice for T.
- The 'more extreme' outcomes will simply be more extreme values of T than what was observed.
- For a one-sided alternate hypothesis, the more extreme values of T are simply the values greater than (or less than) the value of T observed. Greater than or less than depends on the side of H_1. For example, if T is the number of heads, H_0 is $p = 1/2$, and H_1 is $p < 1/2$, then more extreme values of T are values less than what was observed. If H_1 is $p > 1/2$, then more extreme values of T are values greater than what was observed.
- If we draw the CDF for T, we can start to see what to do in the two-sided case.
- Two-sided case: (drawn on board)

2 Continuous CDF's and density functions

What is the relationship between a random variable's density function $f(x)$ and its CDF $F_X(t)$?

Answer: the fundamental theorem of calculus:

$$\begin{aligned}(F_X(t))' &= \left(\int_{-\infty}^{t} f(x)\,dx\right)' \\ &= f(t)\end{aligned}$$

so the density function is the derivative of the CDF.

3 Uses of a CDF

First a simple use of a CDF:

Let X have CDF $F_X(t)$, then

$$\Pr[X \in (a,b]] = F_X(b) - F_X(a)$$

(this is true when X is discrete or continuous)

Next let's see how to answer the following question:

I pick 10 numbers uniformly and independently from $[0,1]$. What's the probability that the maximum of the numbers is at least .9?

We'll answer in steps:

- How to model the situation? Let $X_1, \ldots X_{10}$ be uniform $[0,1]$ rv's, jointly independent.
- Express $\Pr[\max \geq .9]$ as the probability of a *conjuntion* (an *and* of (indpendent) events).
- Write $\Pr[\max \geq .9]$ as a product
- calculate the answer.

Now in general: Let i.i.d. random variables $X_1, \ldots X_n$ all have CDF $F_X(t)$. Calculate:

$$\Pr[\max_i X_i \geq t]$$

4 Dependent continuous RV's

What can we do when X and Y are both continuous rv's but not independent? How do we even specify their dependence?

The answer is a *joint density function*, $f(x,y)$. Properties:

- $f(x,y) \geq 0$
- $\int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y)\, dx\, dy = 1$
- $\Pr[X \in A \cap Y \in B] = \int_B \int_A f(x,y)\, dx\, dy$
- $\Pr[X \in A] = \int_{\mathbb{R}} \int_A f(x,y)\, dx\, dy = \int_A \int_{\mathbb{R}} f(x,y)\, dy\, dx$
- $F_X(t) = \int_{\mathbb{R}} \int_{-\infty}^t f(x,y)\, dx\, dy = \int_{-\infty}^t \int_{\mathbb{R}} f(x,y)\, dy\, dx$
- switching the order of integration uses Fubini's Theorem

5 Questions

1. Let $F(t) = \frac{1}{8}t^3$ for $t \in [0,2]$, $F(t) = 1$ for $t > 2$, 0 for $t < 0$ be a CDF. What is the corresponding density function?
2. For the above RV X, what is $\Pr[X > 1]$?
3. What is $\mathbb{E}X$?
4. Let $Y_1, \ldots Y_n$ be i.i.d. exponential rv's with mean 1. What is $\Pr[\max_i Y_i > 10]$?
5. Using your calculus skills, give the best upper bound you can (in terms of t for large t) for $\Pr[N > t]$ where N is a standard normal r.v., i.e. its density function is $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$