AMS 162

1 Probability Review

Suppose in a class of 20 students, 5 are female, 12 are CS majors, and of these 2 are female CS majors.

$$P(\text{female}) = \frac{5}{20} = \frac{1}{4} \qquad P(\text{CS}) = \frac{12}{20} = \frac{3}{5} \qquad P(\text{female and CS}) = \frac{2}{20} = \frac{1}{10}$$

Recall the rule for computing probabilities of unions (which can be visualized with a Venn diagram):

$$P(\text{female or CS}) = P(\text{female}) + P(\text{CS}) - P(\text{female and CS}) = \frac{5}{20} + \frac{12}{20} - \frac{2}{20} = \frac{3}{4}$$

Conditional probabilities can be thought of as probabilities on a reduced population:

$$P(\text{female}|\text{CS}) = \frac{P(\text{female and CS})}{P(\text{CS})} = \frac{2}{12} = \frac{1}{6}$$

i.e., of the CS students, what fraction of them are female? Similarly, if we consider the female students and ask what fraction of females are CS majors, we get:

$$P(\text{CS}|\text{female}) = \frac{P(\text{CS and female})}{P(\text{female})} = \frac{2}{5}$$

Since $P(\text{female}|\text{CS}) \neq P(\text{female})$, and $P(\text{CS}|\text{female}) \neq P(\text{CS})$, we see that being a CS major is not independent of being female.

2 Inference Review

1. Suppose continuous-valued data are collected, but are recorded by rounding to the nearest integer. A reasonable assumption is that the roundoff error for each number has a uniform distribution on $(-\frac{1}{2}, \frac{1}{2})$. If 100 of these observations are averaged, what is the probability that the answer is correct to within .02, i.e., what is the probability that the average roundoff error is between -0.02 and 0.02?

$$\begin{aligned}
X &\sim U\left(-\frac{1}{2}, \frac{1}{2}\right) \\
E[X] &= 0 \\
Var(X) &= \frac{1}{12} \\
\overline{X} &\mathrel{\dot\sim} N(0, \frac{1}{1200}) \qquad \text{by CLT} \\
P(-.02 < \overline{X} < .02) &= 1 - 2 * P(\overline{X} > .02) = 1 - 2 * P(\frac{\overline{X} - 0}{\sqrt{1/1200}} > \frac{.02}{\sqrt{1/1200}}) \\
&\approx 1 - 2 * P(Z > 0.693) = 1 - 2 * 0.244 = 0.512
\end{aligned}$$

Is the central limit theorem amazing or what?

2. Suppose that at a restaurant, 10 pints of beer are ordered. When the filled glasses arrive, they are carefully measured and it is found that they contain an average of 0.989 pints with a standard deviation of 0.03 pints. Construct a 95% confidence interval for true mean amount of beer in a pint order at this restaurant. What do you conclude about these beers?

$$95\% \text{ CI is } \overline{X} \pm t_{(n-1)0.025}\frac{s}{\sqrt{n}} = 0.989 \pm (2.262)\frac{0.03}{\sqrt{10}} = (0.968, 1.010)$$

   Since 1.0 is in this interval, it is reasonable that this restaurant really does serve 1 pint beers.

3. On one day, there are 80 calls to an IT department's help line. 53 of these were solved on the same day. The company has a policy that 75% of the calls should be resolved on the same day. Is it reasonable that 75% of all calls are resolved on the same day? Also compute a p-value for your test.

   (a) H_0: $p = 0.75$ where p is the population proportion of calls
       H_a: $p \neq 0.75$ resolved on the same day

   (b) $\alpha = 0.05$

   (c) test statistic is $Z = \frac{\hat{p}-p_0}{\sqrt{\hat{p}(1-\hat{p})/n}}$

   (d) by the central limit theorem, if the null hypothesis is true, the test statistic is approximately normally distributed, so we reject if $|Z| > 1.96$

   (e) $\hat{p} = \frac{53}{80} = 0.6625$, so $Z = \frac{0.6625-0.75}{\sqrt{0.6625(1-0.6625)/80}} = -1.655$

   (f) since $-1.655 > -1.96$, we fail to reject H_0 and conclude that it is reasonable that 75% of the calls are resolved on the same day

   (g) the p-value is the probability of getting a test statistic as or more extreme as actually observed, if the null hypothesis is true, i.e., $P(|Z| > |-1.66|) = 2 * P(Z < -1.66) = 2 * (0.0484) = 0.097$. So if it is true that 75% of the calls are resolved on the same day, there is a probability of 0.097 that on a day with 80 calls, either 53 or fewer, or 67 or more calls will be resolved. Thus 53 is not that extreme a number, so there isn't sufficient evidence to dispute the company claim.

3 What is a probability?

What is the probability this box will land upright?

- Is the right answer a fixed number?
- Is there a right answer?
- Does your answer change after observing data? How does it change?

What is the probability that Schwarzenegger is re-elected 2006?

The frequentist or classical perspective views all probabilities as long-run frequencies, and sees them as fixed.

The Bayesian perspective views probabilities as personal, and possibly subjective, and sees them as random. In this perspective, probabilities can also be interpreted as partial information.

In both cases, values for probabilities can be established through a betting interpretation. Your value for the probability p of an event is the amount you would be willing to pay (in \$) for the chance to win \$1 if the event happens. For example, in flipping a fair coin, you should be willing to pay \$0.50 so that if it lands heads you win \$1 (i.e., you lose your \$0.50, but gain an additional \$1 for a net gain of \$0.50) and if it lands tails you lose your \$0.50. Your probabilities must be *coherent* in the sense that they respect the mathematical axioms of probabilities, or else someone can make a "Dutch Book" against you, a series of bets where at the end you are guaranteed to lose money.

3.1 Dutch Book example

Suppose you think that $P(A) = 0.7, P(B) = 0.5$, and that even though A and B are independent, you think that $P(A \text{ and } B) = 0.1$. This means that you think all three of the following are fair bets:

1. Pay 0.70, if A occurs you win \$1, otherwise you win nothing.
2. Pay 0.50, if B occurs you win \$1, otherwise you win nothing.
3. Pay 0.90, if A and B don't both occur you win \$1, otherwise you win nothing.

Since you think all bets are fair, you can take all three at once. If you do, you pay \$2.10, but you only win \$1 if neither A or B happens, and \$2 otherwise. Thus you are guaranteed to lose at least 10 cents, and possibly \$1.10, depending upon the actual probabilities of the events.

Event		Bet			
A	B	1	2	3	Total payout
T	T	T	T	F	2
T	F	T	F	T	2
F	T	F	T	T	2
F	F	F	F	T	1

Any coherent system of probabilities will satisfy the axioms of probability. Note that perception can cause incoherence. People tend to be risk-averse when presented with a situation involving gains, but risk-taking when thinking about avoiding losses.

3.2 Utility Examples

These examples are from "The Framing of Decisions and the Psychology of Choice" by Amos Tversky and Daniel Kahneman, published in *Science*, 1981.

1. Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed, under which:

   A 200 people will be saved

   B with probability 1/3, 600 people will be saved, and with probability 2/3, no people will be saved

2. Under the same scenario as (1), the two alternative programs will result in:

   C 400 people will die

   D with probability 1/3, nobody will die, and with probability 2/3, 600 people will die

3. Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer:

   (i) A a sure gain of $240
   B 25% chance to gain $1000 and a 75% chance to gain nothing

   (ii) C a sure loss of $750
   D 75% chance to lose $1000 and a 25% chance to lose nothing

4. Choose between:

   A & D 25% chance to win $240 and a 75% chance to lose $760

   B & C 25% chance to win $250 and a 75% chance to lose $750

5. Imagine that you have decided to see a play where admission is $10 per ticket. As you enter the theater you discover that you have lost a $10 bill. Would you still pay $10 for a ticket for the play?

6. Imagine that you have decided to see a play and paid the admission price of $10 per ticket. As you enter the theater you discover that you have lost the ticket. The seat was not marked and the ticket cannot be recovered. Would you pay $10 for another ticket?

Notes: When polled, for #1 and 2, 72% of the people preferred A to B, but 78% preferred D to C. For #3, 84% of the people preferred A to B, yet 87% of the people preferred D to C. Of course, when presented with the combinations in #4, everyone realized this was a bad idea. 88% would still buy a ticket in scenario #5, but only 46% would buy another ticket in scenario #6. This line of research was eventually rewarded with a Nobel prize in Economics last year.

4 Inference Example

Your mischievous brother owns a coin which you know to be loaded so that it comes up heads 70% of the time. He comes to you with a coin and wants to make a bet with you. You don't know if this is the loaded coin or a fair one. He lets you flip it 5 times to check it out, and you get 2 heads and 3 tails. Which coin do you think it is? Why? How sure are you?

- What is our *unknown parameter of interest*?
- What is our *probability model* for the data?
- How do we *estimate* our unknown parameter?
- Can we use knowledge from previous experiences?

Probability Model for the Data

Let X = number of heads on five flips. It seems reasonable to model X as a binomial with $n = 5$, and $p = 0.5$ if the coin is fair, $p = 0.7$ if the coin is loaded. Let θ represent the type of coin. Thus:

$$\begin{aligned} P(X = x|\theta = \text{fair}) &= \binom{5}{x}\left(\frac{1}{2}\right)^5 \\ P(X = x|\theta = \text{loaded}) &= \binom{5}{x}(.7)^x(.3)^{5-x} \end{aligned}$$

Learning about θ

Once we observe X (here $X = 2$), we can think of the probability model as a function of θ alone. This function is the *likelihood* and is denoted $L(\theta)$. It gives the probability of the particular observed data as a function of θ.

Maximum Likelihood Estimation

One way to make statements about the value of θ is *Maximum Likelihood*. The idea is that our best guess of the true value of θ is the value that maximizes the likelihood.

$$f(\theta|X = 2) = \begin{cases} 0.3125 & \text{if } \theta = \text{fair} \\ 0.1323 & \text{if } \theta = \text{loaded} \end{cases}$$

So in this example, we would guess that the coin is fair.

Bayesian Approach

A different (and our preferred) approach to making statements about the value of θ is the *Bayesian* approach. In this case we think of θ as a random variable. We start with a *prior* distribution for θ, $P(\theta)$, (representing our beliefs before seeing the data). We then observe the data and compute a *posterior* distribution, $P(\theta|X)$, by using Bayes' Theorem.

$$P(\theta|X) = \frac{P(\theta, X)}{P(X)} = \frac{P(X|\theta)P(\theta)}{\int P(X|\theta)P(\theta)d\theta} \propto L(\theta)P(\theta)$$

Example continued

Since you know your brother reasonably well, you might think that there is a 60% chance that this is the loaded coin, before you get to flip it 5 times. You then get 2 heads and 3 tails. What is your posterior probability that the coin is loaded?

$$Pr(\text{loaded}|X) = \frac{Pr(X|\text{loaded})Pr(\text{loaded})}{Pr(X|\text{loaded})Pr(\text{loaded}) + Pr(X|\text{fair})Pr(\text{fair})}$$

$$= \frac{\binom{5}{2}(.7)^2(.3)^3(.6)}{\binom{5}{2}(.7)^2(.3)^3(.6) + \binom{5}{2}(.5)^2(.5)^3(.4)} = \frac{0.07938}{0.07938 + 0.125} = 0.388$$

Prior Sensitivity

We could have made a different choice for a prior:

- $Pr(\text{loaded}) = .9 \Rightarrow Pr(\text{loaded}|X) = 0.792$
- $Pr(\text{loaded}) = .5 \Rightarrow Pr(\text{loaded}|X) = 0.297$

Note that the "non-informative" prior gives the "same" answer as the MLE. Also note that since the sample size is small, the result is highly sensitive to the choice of prior. As the sample size increases, the prior matters less.

Additional Benefits

By thinking of the parameter(s) as random, we gain a number of additional benefits in estimating uncertainty about the parameter and in interpreting our results.

Intervals

With a standard confidence interval, we can say "we are 95% confident that the true value of the parameter is in this interval", but this is a code phrase to evade the question as to whether or not we think the true value really is in the interval or not. We can't attach any real probability to it. However under the Bayesian approach, if we make a posterior interval (usually called a credible interval), then we can say that there is a 95% probability that the parameter is in the interval.

Hypothesis Testing

In classical statistics, we treat the two hypotheses differently, and can only show that the null is false. We can never prove that the null is really true. However as a Bayesian, one can put prior probabilities that each of the hypotheses is true, and then learn from the data and compute the posterior probabilities that the hypotheses are true.