Math 455 Class 38 April 29

More about midterm 2

Problem 5:

The first step is to use the equation

$$Y = 2 + X_1 + 3X_2 + \varepsilon$$

Notice that the mis-specified model includes an ε'. You are not told anything about ε'. In particular, you do not know that $\mathrm{Cov}(X_1, \varepsilon') = 0$.

After the first step, a common error was to write

$\mathrm{Cov}(X_2, X_1) =$ something. not right.

We know: $\operatorname{Cor}(X_1, X_2) = \rho \quad \operatorname{Var}(X_1) = \sigma_1^2$

$$\operatorname{Var}(X_2) = \sigma_2^2.$$

Which formula for $\operatorname{Cov}(X_1, X_2)$ in terms of $\operatorname{Cor}(X_1, X_2)$ is correct?

- $\operatorname{Cov}(X_1, X_2) = \rho$
- $\operatorname{Cov}(X_1, X_2) = \rho\, \sigma_1^2 \sigma_2^2$
- $\operatorname{Cov}(X_1, X_2) = \rho \sqrt{\sigma_1^2 + \sigma_2^2}$
- $\operatorname{Cov}(X_1, X_2) = \dfrac{\rho}{\sigma_1 \sigma_2}$

Correct!

- $\operatorname{Cov}(X_1, X_2) = \rho\, \sigma_1 \sigma_2$
- $\operatorname{Cov}(X_1, X_2) = \rho^2 \sigma_1^2 \sigma_2^2$

Note analogy with

$$\langle v, w \rangle = \|v\| \, \|w\| \cos \theta$$

A model of the stock market

$X_1, \ldots, X_N$: % return on stock i over the next year

$\text{Var}(X_i) = \sigma^2$: all stocks have the same volatility.

$\text{Cor}(X_i, X_j) = \rho$: for all $i \neq j$.

Think : $N = 1000$ $\sigma = 0.25 = 25\%$
$\rho = 0.2$

⚠ of course the real stock market is more complex.

Two portfolio managers : P, Q.

$$P = \frac{1}{n}(X_{i_1} + \cdots + X_{i_n})$$ ← return on portfolio P consisting of $X_{i_1}, \ldots, X_{i_n}$

$$Q = \frac{1}{n}(X_{j_1} + \cdots + X_{j_n}).$$ ← return on Q.

Think : $n = 100$. P, Q have no stocks in common.

Q: what is the correlation of the return on P and the return on Q?

You might think: the correlation ρ of two different stocks is low, and P, Q have nothing in common, so $\mathrm{Cor}(P, Q)$ is low.

This is WRONG.

A calculation at the Math 447 level shows that the correlation is high.

$$\mathrm{Cov}(P, Q) = \mathrm{Cov}\left(\frac{1}{n}\sum_{\ell=1}^{n} X_{i_\ell}, \frac{1}{n}\sum_{m=1}^{n} X_{j_m}\right)$$

$$= \frac{1}{n^2}\sum_{\ell}\sum_{m} \mathrm{Cov}(X_{i_\ell}, X_{j_m})$$

$$= \frac{1}{n^2}\cdot n^2 \cdot \rho\sigma^2 = \rho\sigma^2$$

$$\mathrm{Cor}(P, Q) = \frac{1}{\sigma_P \sigma_Q}\mathrm{Cov}(P, Q)$$

$$\sigma_P = \sqrt{\mathrm{Cov}(P, P)}$$

$$\sigma_P^2 = \frac{1}{n^2} \sum_{l} \sum_{m} \mathrm{Cov}(X_{i_{l(\neq)}}, X_{i_m})$$

$$= \frac{1}{n^2} \left(\sum_{l \neq m} \rho\sigma^2 + \sum_{l=m} \sigma^2 \right)$$

$$= \frac{1}{n^2} \left[(n^2 - n)\rho\sigma^2 + n\sigma^2 \right]$$

$$= \sigma^2 \left(\frac{1}{n} + \rho \left(1 - \frac{1}{n}\right) \right)$$

$\sigma_Q^2 = \sigma_P^2$, so $\sigma_P \sigma_Q = \sigma_P^2$.

and $\mathrm{Cor}(P, Q) = \dfrac{n\rho}{1 + (n-1)\rho}$

With $n = 100$, $\rho = 0.2$, this is ≈ 0.96

High, not low.

This idea has meaningful consequences:

- Much of the performance of a portfolio of stocks is explained by the performance of the stock market as a whole.

- The Capital Asset Pricing Model

  (Sharpe, Nobel 1990) (Also Treynor, Linter, Mossin)

- The "index fund": we don't need to pay a

  fund manager, we just buy the "index",

  own all stocks in proportion to market capitalization.

Math 532 May 12.

Bayesians, Frequentists, Likelihood Principle

The most important thing to know: in most applications, nobody cares.

The ultimate validation in business: We made money.

Arguing philosophy of statistics does not make money.

So take the following with a grain of salt.

An example of Savage:

We are testing a coin with probability of heads θ for bias. We consider

$$H_0: \theta = 1/2 \qquad \text{vs.} \qquad H_1: \theta > \frac{1}{2}.$$

Experiment A: We flip the coin 12 times and get 9 heads.

X., the number of heads is $X \sim Bin(12, \theta)$

and $$P(X \geq 9 \mid H_0) = \sum_{x=9}^{12} \binom{12}{x} \left(\frac{1}{2}\right)^x \left(1-\frac{1}{2}\right)^{12-x} \approx 0.073.$$

So H_0 is NOT rejected at $\alpha = 0.05$

Experiment B: We flip the coin until 3 tails are observed and count X, the number of heads.

We observe $X = 9$.

$X \sim NB(3, 1-\theta)$ (negative binomial)

$$P(X \geq 9 \mid H_0) = \sum_{x=9}^{\infty} \binom{3+x-1}{2} \left(\tfrac{1}{2}\right)^x \left(\tfrac{1}{2}\right)^3$$

$$\approx 0.033$$

So H_0 is rejected at $\alpha = 0.05$.

OK, fine: these are different experiments.

Different rules governing collection of data.

However: in both cases we observed 9 heads and 3 tails.

The Likelihood Principle: after we observe the data, all relevant information is contained in the likelihood function.

What are the likelihood functions in A and B?

$$A: \quad f(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} = \binom{12}{9} \theta^9 (1-\theta)^3$$

$$= 220\, \theta^9 (1-\theta)^3$$

$$B: \quad f(x \mid \theta) = \binom{\alpha + x - 1}{\alpha - 1} (1-\theta)^\alpha \theta^x = \binom{3+9-1}{3-1} (1-\theta)^3 \theta^9$$

$$= 55\, \theta^9 (1-\theta)^3.$$

Note that multiplying the likelihood function by a constant (here 4) does not change anything.

If we are strictly Bayesian and accept the likelihood principle, then the evidence of experiments A and B is the same.

In other words, the rules governing when data collection stops are not relevant to interpretation of the data.

Many people have a hard time accepting this.

The Chow-Robbins Game.

We flip a fair coin repeatedly, and can stop whenever we like. If, upon stopping, we have h heads in n ~~fl~~ flips, we are paid $\frac{h}{n}$ dollars.

Example 1: We flip THHH and elect to stop: we are paid 75¢.

Example 2: We flip H and elect to stop. We are paid \$1.

Example 3: We flip ~~H~~ T T and elect to stop (bad strategy): we are paid \$0.

The exact value of the game, and a perfect strategy, is unknown.

Partial results, based on the law of the iterated logarithm, are known.

Easy proof that the value of the game is at least 75¢:

Simple strategy: If the first flip is H, stop. otherwise flip 1000 times.

We get about $75¢ = \frac{1}{2}\cdot(\$1) + \frac{1}{2}\cdot(50¢)$.

To see the difficulty, show that it is correct with 2 H and 1 T to continue but with 5H and 3T to stop.

Now what if we have 66H and 59T?

This gives most people the intuition that the rules of data collection do matter.

Again: in the workplace none of this matters.

Very simple ideas can be a major contribution.

I will post a link to a NY Times article about statistics in ~~the~~ advertising.