140.651 - Lab 8

1. Law of large numbers

From Lecture 8 (slides 10-20):

If $X_1, ..., X_n$ are iid from a population with mean μ and variance σ^2 then $\bar{X}_n$ converges to μ in probability.

2010 Final Problem 1

1. You simulate 10 variables from a normal distribution with mean 0 and variance 1 and 100 more from a normal distribution with mean 5 and variance 1. You repeat this process (simulating a total of 110 normals) $I = 10,000$ times. Let S_{1i}^2 and S_{2i}^2 and $\bar{X}_{1i}$ and $\bar{X}_{2i}$ be the sample means and variances for sample $i = 1, \ldots, I$, respectively. Answer the following (it is not necessary to solve for final decimal numbers):

   A. About what number will $\frac{1}{I}\sum_{i=1}^{I}(S_{1i}^2 + S_{2i}^2)$ be close to?

   B. Let $D_i = \bar{X}_{2i} - \bar{X}_{1i}$. About what number will $\bar{D} = \frac{1}{I}\sum_{i=1}^{I} D_i$ be close to?

   C. About what number will $\frac{1}{I-1}\sum_{i=1}^{I}(D_i - \bar{D})^2$ be close to?

(A) About what number will $\frac{1}{I}\sum_{i=1}^{I}(S_{1i}^2 + S_{2i}^2)$ be close to?

- Use the Law of Large Numbers. View $\left(S_{1i}^2 + S_{2i}^2\right)$, $i = 1, \ldots, I$ as I individual observations.
- As I goes to infinity, $\frac{1}{I}\sum_{i=1}^{I}(S_{1i}^2 + S_{2i}^2)$ converges in probability to $E[S_{11}^2 + S_{21}^2]$.
- S_{11}^2 and S_{21}^2 are unbiased estimators of the variance of the two sample: $E[S_{11}^2] = E[S_{21}^2] = 1$ and $E[S_{11}^2 + S_{21}^2] = 2$. Hence $\frac{1}{I}\sum_{i=1}^{I}(S_{1i}^2 + S_{2i}^2)$ will be close to 2.

(B) Let $D_i = \bar{X}_{2i} - \bar{X}_{1i}$. About what number will $\bar{D} = \frac{1}{I}\sum_{i=1}^{I} D_i$ will be close to?

Lecture material revisited:

- Sample means of normal random variables are normally distributed.
- The difference of two normal random variables are normally distributed.

$$\bar{X}_1 \sim N(0, 1/10), \bar{X}_2 \sim N(5, 1/100).$$

.

$$E[D_i] = E[\bar{X}_2] - E[\bar{X}_1] = 5.$$

Since $\bar{X}_1$ and $\bar{X}_2$ are independent, $var(D_i) = var(\bar{X}_1) + var(\bar{X}_2) = 11/100$.

Therefore $D_i \sim N(5, 11/100)$.

Again by the LLN, $\bar{D}$ converges to 5 in probability.

(C) About twhat number will $\frac{1}{I-1}\sum_{i=1}^{I}(D_1 - \bar{D})$ be close to?

$$\frac{1}{I-1}\sum_{i=1}^{I}(D_i - \bar{D}) \xrightarrow{p} var(D_1) = 11/100.$$

See lecture 8 page 18 for details.

2. Likelihood

- The likelihood of the data is the joint density evaluated as a **function of the parameters** with the data fixed.
  - Given data $\boldsymbol{x}$ and PMF or PDF $f(\boldsymbol{x}, \theta)$, the likelihood is $L(\theta \mid \boldsymbol{x}) = f(\boldsymbol{x}, \theta)$.
- Ratios of likelihood values measure the relative evidence of one value of the unknown parameter to another.

2010 Final Problem 2

2. You glue together a quarter, nickel, penny and dime (in that order) to obtain a funny shaped coin with a head on the small side and a tail on larger one. You claim that the coin is fair while a friend claims that it should have probability of a head of 25%. Your friend flips the coin 5 times to obtain 2 heads and 3 tails. Write out a number that would compare the relative evidence of the two hypotheses. (You do not need to calculate the final number, simply plug into the relevant equations.)

Compute the likelihood ratio!

Denote by p the probability of heads.

Likelihood $L(p \mid \boldsymbol{x}) = \frac{5!}{2!3!}p^2(1-p)^3$.

Likelihood ratio $\frac{L(0.5|\boldsymbol{x})}{L(0.25|\boldsymbol{x})} = \frac{\frac{5!}{2!3!}0.5^2(1-0.5)^3}{\frac{5!}{2!3!}0.25^2(1-0.25)^3} \approx 1.19$. There is more evidence supporting that the coin is fair than that the probability of heads is 0.25.

3. Maximum likelihood

Three steps to compute the maximum likelhood estimate:

1. Write down the log-likelihood $l(\theta \mid \boldsymbol{x}) = logL(\theta \mid \boldsymbol{x}) = \sum_{i=1}^{n} logf(x_i \mid \theta)$.
2. Solve the score equation $\frac{d}{d\theta}l(\theta \mid \boldsymbol{x}) = 0$.
3. Check $\frac{d^2}{d\theta^2}l(\theta \mid \boldsymbol{x}) < 0$.

2010 Final Problem 3

3. The Poisson mass function is for a random count of events for a process having been monitored for a fixed (non-random) time t is given by:

$$\frac{(\lambda t)^x \exp(-\lambda t)}{x!} \quad \text{for} \quad x = 0, 1, \ldots.$$

Suppose that $x_1, \ldots, x_N$ are independent counts of events with associated monitoring times $t_1, \ldots, t_N$. Argue that the maximum likelihood estimate of λ is

$$\hat{\lambda} = \frac{\sum_{i=1}^N x_i}{\sum_{i=1}^N t_i}.$$

1. Since the observations are independent, the likelihood is the product of individual PMFs:

$$L(\lambda \mid \boldsymbol{x}) = \prod_{i=1}^N \frac{(\lambda t_i)^{x_i} \exp(-\lambda t_i)}{x_i!} = \lambda^{\sum_{i=1}^N x_i} \exp\Big(- \lambda \sum_{i=1}^N t_i\Big) \prod_{i=1}^N \frac{t_i^{x_i}}{x_i!}$$

The log-likelihood is

$$l(\lambda \mid \boldsymbol{x}) = (\sum_{i=1}^N x_i) log(\lambda) - \lambda \sum_{i=1}^N t_i + C.$$

C is a constant.

2. The score equation is

$$\frac{d}{d\lambda} l(\theta \mid \boldsymbol{x}) = \frac{\sum_{i=1}^N x_i}{\lambda} - \sum_{i=1}^N t_i = 0.$$

Soving the equation gives

$$\hat{\lambda} = \frac{\sum_{i=1}^N x_i}{\sum_{i=1}^N t_i}.$$

3. Compute the second derivative

$$\frac{d^2}{d\lambda^2} l(\lambda \mid \boldsymbol{x}) = -\frac{\sum_{i=1}^N x_i}{\lambda^2} < 0.$$

Therefore $\hat{\lambda} = \frac{\sum_{i=1}^N x_i}{\sum_{i=1}^N t_i}$ is the MLE of λ.

4. Confidence intervals

2010 Final Problem 6

6. A friend is study hypertension and wants to estimate the prevalence (percentage of people) having hypertension in a specific population using a 95% Wald interval on a sample of n subjects

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

How large would n have to be to have the margin of error (1/2 the width of the confidence interval) no larger than .01 regardless of the value of $\hat{p}$.

1. (Derivation) The sample prevalence $\hat{p}$ is a consistent (and unbiased) estimate of the population prevalence p.

$$E[\hat{p}] = p, var(\hat{p}) = p(1-p)/n$$

.

Using the Central Limit Theorem, when the sample size is large, $\hat{p}$ is approximately distributed as $N\left(p, \frac{p(1-p)}{n}\right)$.

Hence

$$P\left(|\hat{p} - p| \leq 2\sqrt{p(1-p)/n}\right) = 0.95 = P\left(\hat{p} - 2\sqrt{p(1-p)/n} \leq p \leq \hat{p} + 2\sqrt{p(1-p)/n}\right).$$

Hence the confidence interval is $\hat{p} \pm 2\sqrt{p(1-p)/n}$.

Replace p by $\hat{p}$ gives CI $\hat{p} \pm 2\sqrt{\hat{p}(1-\hat{p})/n}$. (Why is this substitution correct? Hint: Slutsky's theorem.)

2. (Size of CI) Half the length of the confidence interval is

$$2\sqrt{\hat{p}(1-\hat{p})/n} \leq 2\sqrt{1/(4n)} = 1/\sqrt{n} \leq 0.01$$

.

$$n \geq 10,000.$$

Review:

- Confidence intervals are random! The parameter is fixed.
- For example, if we obtain a 95% percent confidence interval for [0.45,0.57]
  - It is **wrong** to interpret it as "the probability for p to be in [0.45,0.57] is 95%".
  - The correct interpretation is: if we repeat the same procedure many times and get a confidence interval each time, 95% of those intervals will include p. One of the realizations of the interval is [0.45,0.57].
- The length of the confidence interval is proportional to $1/\sqrt{n}$.