# STA261: Partial Solutions, Assignments 1 - 5

*Alex Stringer*

*February, 2018*

These are fully explained solutions to select assignment questions from assignments 1 - 5. Use these to verify your answers, and to guide your answers to similar problems.

**Assignment 1, Question 5**: Let $\{X_n\}$ be a sequence of random variables with $E(X_i) = \mu$ and $\lim_{n\to\infty} Var(X_n) = 0$. Show $X_n \xrightarrow{p} \mu$.

*Solution*: recall Chebyshev's inequality: for random variable $X$ with $E(X) = \mu$ and $Var(X) = \sigma^2$, for any $t > 0$ we have

$$P(|X - \mu| > t) \leq \frac{\sigma^2}{t}$$

As stated in the question, $X_n$ has $E(X_n) = \mu$ and $Var(X_n) = \sigma_n^2$. Since probabilities are bounded below by 0, if $\sigma_n^2 \to 0$ as $n \to \infty$ then using Chebyshev we have

$$0 \leq P(|X_n - \mu| > t) \leq \frac{\sigma_n^2}{t} \to 0$$

Because for any $t > 0$, $P(|X_n - \mu| > t)$ is bounded below and above by 0 in the limit, by the so-called "squeeze theorem" from calculus,

$$\lim_{n\to\infty} P(|X_n - \mu| > t) = 0$$

so by definition, $X_n \xrightarrow{p} \mu$.

**Assignment 1, Question 7**: Let $\{X_i\}$ be a sequence of independent random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$.

(a) Evaluate $\lim_{n\to\infty} P(|\bar{X_n} - \mu| > 0.01)$.
(b) Can you evaluate $P(|\bar{X_{100}} - \mu| > 0.01)$? Why or why not?

*Solution*:

(a) This equals 0 by the definition of the Law of Large Numbers, which states that $\lim_{n\to\infty} P(|\bar{X_n} - \mu| > \epsilon) = 0$ for any $\epsilon > 0$.
(b) No. The question did not give you any information about the distribution of the $X_i$. You could argue that you could *approximate* that probability using the Central Limit Theorem, although that's not what the question asks and you weren't given the variance $\sigma^2$, which you need.

**Assignment 1, Question 8**: Let $\{X_i\}$ be a sequence of independent random variables having the discrete uniform distribution, $X_i \sim unif\{-1, 1\}$.

(a) Evaluate $E(X_i)$ and $Var(X_i)$

(b) Use the central limit theorem to calculate

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{\sqrt{n}}\right| \leq 1\right)$$

(c) Derive the exact distribution of $\sum_{i=1}^n X_i$. *Hint*: find a simple transformation that makes each $X_i$ Bernoulli. Then the distribution of the sum is known.

*Solution*:

(a)

$$\begin{aligned}
E(X_i) &= \sum_x xP(X=x) = (-1)\times(1/2) + (1)\times(1/2) = 0 \\
E(X_i^2) &= \sum_x x^2 P(X=x) = (-1)^2\times(1/2) + (1)^2\times(1/2) = 1 \\
Var(X_i) &= E(X_i^2) - E(X_i)^2 = 1
\end{aligned}$$

(b) You need

$$\begin{aligned}
E\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n E(X_i) = 0 \\
Var\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n Var(X_i) = n
\end{aligned}$$

The central limit theorem then says that

$$\frac{\sum_{i=1}^n X_i}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

By the definition of convergence in distribution, we can then write

$$\begin{aligned}
\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{\sqrt{n}}\right| \leq 1\right) &= P\left(|Z| \leq 1\right) \\
&= P(-1 \leq Z \leq 1) \\
&= \Phi(1) - \Phi(-1) \\
&\approx 0.683
\end{aligned}$$

where $Z \sim N(0,1)$ and $\Phi$ is the standard normal $CDF$. You can use a normal table or a computer to compute the numerical answer. Note that on a test, I am fine with the answer being left in terms of $\Phi$.

(c) Let $Y_i = \frac{1}{2}(X_i + 1)$. Then $Y_i$ is a Bernoulli random variable with $p = 1/2$, because it takes on values 0 or 1 each with probability 1/2. Therefore,

$$\sum_{i=1}^n Y_i = \sum_{i=1}^n \frac{1}{2}(X_i+1) \sim Binom(n, 1/2)$$

You want the distribution of $\sum_{i=1}^{n} X_i$. When asked for a distribution, you can give a probability density/mass function, cumulative distribution function, or moment generating function. Try them all if you like; in this case, the CDF is

$$\begin{aligned} P\left(\sum_{i=1}^{n} X_i \leq x\right) &= P\left(\sum_{i=1}^{n} Y_i \leq \frac{1}{2}(x+1)\right) \\ &\equiv F\left(\frac{1}{2}(x+1)\right) \end{aligned}$$

where $F(\cdot)$ is the CDF of a $Binom(n, 1/2)$ random variable. It is a messy answer, but it is still an answer, and it is enough to do parts d and e if you so choose.

**Assignment 2, Question 1**: In the following problems, identify the *parameter*, the *estimator*, and the *estimate*. There might be 0, 1, or 2 of each thing in each question.

(a) We sample the heights of U of T students, which we know to have a mean of $170cm$. We get a sample mean of $168cm$.

(b) An auto insurance company knows that the probability of someone with your age and driving record making a claim in the next 30 days is 0.0001. They count the number of claims you make in this period, and find you made 0.

(c) An auto insurance company wants to figure out the expected number of claims that someone with your age and driving record should make in the next 30 days. They look at similar 30 day periods, and calculate that out of 40,000 such individuals, 37 claims were made.

(d) The news reports that for a poll in an election campaign, candidate A has a popularity of 49% and candidate B has a popularity of 51%, and that this means candidate B is going to get the most votes in the election.

*Solution*: This question was designed to be vague. If you get the difference between the *parameter* (fixed, unknown value we want to estimate), the *estimator* (function which takes in the data and returns a number/vector of numbers) and an *estimate* (the number/vector returned by the estimator, that we use as our best guess of the parameter) then you are good to go.

(a) "Which we know to have a mean of 170cm"- 170cm is the "true mean" in this question, therefore it is the parameter. The estimate is 168cm, and "the sample mean" would be the estimator.

(b) The parameter to be estimated here is the true probability of someone with your age and driving record making an insurance claim. There is no estimator or estimate explicitly given, but some students suggested that 0 is a reasonable estimate, given the information in the question. There is nothing wrong with that answer.

(c) Now, they take a sample of similar people. The parameter is the same as in (b), but now you may suggest that a reasonable estimator is the sample proportion and the estimate is 37/40,000.

(d) There are several different ways to identify the parameter(s): the single parameter could be the popularity of candidate A, or B; or there could be two parameters (popularity of both candidates), but they have to add to 1. The estimator is not explicitly given, but if you thought a sample mean was appropriate you wouldn't be wrong. The estimate(s) given is/are 49% and 51%.

**Assignment 2, Question 2**: For independent random samples from the following families of distributions, show that the given estimator is consistent for the population parameter.

(a) $X_i \sim Pois(\lambda)$, $\hat{\lambda} = \bar{X}$.

(b) $X_i \sim Exp(\theta)$, where $f_\theta(x) = \theta e^{-\theta x}$, $\hat{\theta} = 1/\bar{X}$.

(c) $X_i \sim Exp(\beta)$, where $f_\beta(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$, $\hat{\beta} = \bar{X}$

(d) $X_i \sim \chi^2_\nu$, $\hat{\nu} = \bar{X}$

(e) $X_i \sim Gamma(\alpha, \beta)$ with $f_{\alpha,\beta}(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$, $\hat{\alpha} = \frac{(\bar{X})^2}{s^2}$ and $\hat{\beta} = \frac{s^2}{\bar{X}}$, where $s^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$. Note you can use $\frac{1}{n-1}$ in the sample variance if you want- they both give a consistent estimator of $\sigma^2$.

*Solution*: For each of these, you have to compute the moment(s) as a function of the parameter(s), then use the law of large numbers.

(a) $E(X) = \lambda$, so $\hat{\lambda} = \bar{X} \xrightarrow{p} E(X) = \lambda$ by the LLN, and is consistent.

(b) $E(X) = 1/\theta$, so $\hat{\theta} = 1/\bar{X} \xrightarrow{p} 1/E(X) = \theta$ by the LLN and the fact that $f(x) = 1/x$ is continuous. So $\hat{\theta}$ is consistent.

(c) $E(X) = \beta$ so $\hat{\beta} = \bar{X} \xrightarrow{p} E(X) = \beta$ by the LLN, and is consistent.

(d) If you hadn't seen the $\chi^2$ distribution before, don't worry, you could skip this one. The answer is the same as the above though; $E(X) = \nu$, etc.

(e) This was the only complicated one. With the given parameterization, you can show that

$$E(X) = \alpha\beta$$
$$Var(X) = \alpha\beta^2$$

which gives

$$\alpha = \frac{E(X)^2}{Var(X)}$$
$$\beta = \frac{Var(X)}{E(X)}$$

Then because the LLN says $\bar{X} \xrightarrow{p} E(X)$, $s^2 \xrightarrow{p} Var(X)$, and all the functions involved are continuous,

$$\hat{\alpha} \xrightarrow{p} \alpha$$
$$\hat{\beta} \xrightarrow{p} \beta$$

**Assignment 2, Question 3**: For $X_i \sim N(\mu, \sigma)$, state and prove whether each estimator is consistent or not. Be sure to say exactly where you are assuming a function is continuous in your proofs.

(a) $\hat{\mu} = \bar{X}$

(b) $\hat{\sigma} = s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$

(c) $\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$

(d) $\hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}\left|X_i - \bar{X}\right|$

(e) $\hat{\mu} = \frac{1}{n+1{,}000{,}000}\sum_{i=1}^{n} X_i$

(f) $\hat{\mu} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_i}}$

(g) $\hat{\sigma^2} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_i^2} - \left(\frac{1}{n}\sum_{i=1}\frac{1}{X_i}\right)^2}$

*Solution*: the point of this question is to get you using the law of large numbers correctly. The LLN implies that

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} E(X)$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{p} E(X^2)$$

$$\frac{1}{n}\sum_{i=1}^{n}\left|X_i - \bar{X}\right| \xrightarrow{p} E(\left|X - E(X)\right|)$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_i} \xrightarrow{p} E\left(\frac{1}{X}\right)$$

and so on. Use these facts, combined with the examples done in lecture, to deduce that

(a) Consistent

(b) Consistent

(c) Consistent

(d) Not Consistent

(e) Consistent

(f) Not Consistent

(g) Not Consistent

The only one that doesn't use this logic is (e); this is consistent because $\frac{1}{n+1{,}000{,}000}\sum_{i=1}^{n} X_i$ must have the same limit as $n \to \infty$ as $\frac{1}{n}\sum_{i=1}^{n} X_i$.

**Assignment 2, Question 4**: Recall the covariance between two random variables $X$ and $Y$ is defined as

$$Cov(X,Y) = E((X - E(X))(Y - E(Y)))$$

(a) Show $Cov(X,Y) = E(XY) - E(X)E(Y)$

(b) Using what you know about convergence in probability of continuous functions of random variables, show that the *sample correlation coefficient*

$$R = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

is consistent for the population correlation,

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

*Solution*:

(a)

$$\begin{aligned}
Cov(X,Y) &= E((X - E(X))(Y - E(Y))) \\
&= E(XY - XE(Y) - YE(X) + E(X)E(Y)) \\
&= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\
&= E(XY) - E(X)E(Y)
\end{aligned}$$

This works because $E(X)$ and $E(Y)$ are just constants (numbers), so $E(XE(Y)) = E(Y)E(X)$ and so on.

(b) Divide both the top and bottom by $n$, and then note that by the LLN, the fact that all the functions are continuous, and multiple applications of Slutky's lemmas, we find that

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \xrightarrow{p} E((X - E(X))(Y - E(Y)))$$
$$\frac{1}{n}\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2} \xrightarrow{p} \sqrt{Var(X)Var(Y)}$$

And applying these facts (continuous functions and Slutsky's lemmas) again,

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \xrightarrow{p} \frac{E((X - E(X))(Y - E(Y)))}{\sqrt{Var(X)Var(Y)}}$$

**Assignment 2, Question 7**: Let $\epsilon_i \sim N(0,\sigma)$ be independent and identically distributed random variables. Let $\beta \in \mathbb{R}$ be a fixed, unknown constant and $x_i \in \mathbb{R}, i = 1 \ldots n$ be fixed, *known* quantities. Let $Y_i = \beta x_i + \epsilon_i, i = 1 \ldots n$

(a) Identify all the *parameters* in this question. There are two.
(b) Find a Method of Moments estimator for $\beta$.
(c) Even though I told you MoM estimators are always consistent- prove that your estimator for $\beta$ is consistent by going through the usual motions.
(d) Find a Method of Moments estimator of $\sigma^2$, and show that it too is consistent.

*Solution*: Note that $Y_i \sim N(\beta x_i, \sigma)$, so $E(Y_i) = \beta x_i$. Take the sample mean of both sides for $i = 1 \ldots n$,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} Y_i &= \beta \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n}\epsilon_i \\
&= \beta\bar{x} + \bar{\epsilon}
\end{aligned}$$

By applying the LLN to $\bar{\epsilon}$ and Slutsky's lemma for addition,

$$\bar{Y} \xrightarrow{p} \beta\bar{x} + E(\hat{\epsilon}) = \beta\bar{x}$$

Hence $\bar{Y}$ is a consistent estimator for $\beta\bar{x}$. Since the $x_i$ are fixed constants, you can just rearrange to obtain

$$\hat{\beta} = \frac{\bar{y}}{\bar{x}}$$

This is an example of a *statistical model*: we took a probability distribution, and wrote its mean as a function of some parameters and input variables. Some students came up with other arguments for solving this problem that led to the same estimator, which is great. Typically, though, MoM wouldn't be used in a confusing example like this; if you try to get the MLE, it will be a lot easier.

**Assignment 3, Question 2**: Show the following estimators are sufficient for their respective population parameters, for the following independent random samples and corresponding distributions. If you use the *factorization theorem*, be sure to state the functions $g(\hat{\theta}, \theta)$ and $h(\mathbf{x})$.

(a) $X_i \sim Gamma(\alpha, \beta)$ with density $f_{x_i}(x_i) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x_i^{\alpha-1} e^{-\frac{x_i}{\beta}}$, $(\hat{\alpha}, \hat{\beta}) = \left(\prod_{i=1}^n x_i, \sum_{i=1}^n x_i\right)$

(b) $X_i \sim N(\mu, \sigma)$, $(\hat{\mu}, \hat{\sigma}) = \left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2\right)$

(c) $X_i \sim N(\mu, \sigma)$, $(\hat{\mu}, \hat{\sigma}) = \left(\bar{x}, \bar{x^2}\right)$

(d) $X_i \sim Beta(\alpha, \beta)$ with $f_{x_i} = \frac{\Gamma(\alpha+\beta)}{\Gamma\alpha\Gamma\beta} x_i^{\alpha-1} (1 - x_i)^{\beta-1}$, $\left(\hat{\alpha}, \hat{\beta}\right) = \left(\prod_{i=1}^n x_i, \prod_{i=1}^n (1 - x_i)\right)$

(e) $X_i \sim Beta(\alpha, \beta)$ as before, $\left(\hat{\alpha}, \hat{\beta}\right) = \left(\sum_{i=1}^n \log x_i, \sum_{i=1}^n \log(1 - x_i)\right)$

*Solution*:

(a) Joint density is

$$f(\mathbf{x}|\alpha, \beta) = \left(\frac{1}{\Gamma(\alpha)\beta^\alpha}\right)^n \left(\prod_{i=1}^n x_i^{\alpha-1}\right) e^{-\frac{\sum_{i=1}^n x_i}{\beta}}$$

which factors with

$$g(\hat{\alpha}, \hat{\beta}, \alpha, \beta) = f(\mathbf{x}|\alpha, \beta)$$
$$h(\mathbf{x}) = 1$$

d) Joint density is

$$f(\mathbf{x}|\alpha, \beta) = \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)^n \left(\prod_{i=1}^n x_i^{\alpha-1}\right) \left(\prod_{i=1}^n (1 - x_i)^{\beta-1}\right)$$

which factors with

$$g(\hat{\alpha}, \hat{\beta}, \alpha, \beta) = f(\mathbf{x}|\alpha, \beta)$$
$$h(\mathbf{x}) = 1$$

For (c) and (e), note that the statistics given are one-to-one functions of the sufficient statistics for those distributions given in (b) and (d).

**Assignment 3, Question 3**: For $X_i \sim Unif(a, b)$, the *continuous* uniform distribution on $(a, b)$, find a sufficient statistic for $(a, b)$. *Hint*: the density is only defined over a certain subset of $\mathbb{R}$, what is it? Make sure to include the corresponding *indicator function* of the support when you write out the density, i.e.

$$f_{x_i}(x_i) = \frac{1}{b - a} \times I(support)$$

It's actually good form to *always* do this, even if I often don't do it for you.

*Solution*: the joint density is

$$f(\mathbf{x}|a,b) = \left(\frac{1}{b-a}\right)^n \times \prod_{i=1}^n I\left(a \leq x_i \leq b\right)$$

Focus on the product of indicator functions: note that the whole thing equals 1 if all the $x_i$ are between $a$ and $b$, and the whole thing equals 0 if any of the $x_i$ are not between $a$ and $b$. So you may write

$$\prod_{i=1}^n I\left(a \leq x_i \leq b\right) = I(a \leq \min_{i=1}^n x_i \leq \max_{i=1}^n x_i \leq b)$$

The result then follows by the factorization theorem with

$$g(\hat{a}, \hat{b}, a, b) = f(\mathbf{x}|a,b)$$
$$h(\mathbf{x}) = 1$$

**Assignment 3, Question 8**: Let $X_i \sim Bern(p)$ be a sequence of independent coin flips. Find the likelihood function for $\mathbf{x} = (x_1, \ldots, x_n)$. Compare your answer to the *binomial* probability mass function, and explain why they are slightly different.

*Solution*: the pmf for each $x_i$ is

$$P(X_i = x_i|p) = p^{x_i}(1-p)^{1-x_i}$$

This is an independent sample, so the likelihood function, which is the joint density of the sample, is obtained by multiplying the marginals:

$$L(p) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} \times (1-p)^{n - \sum_{i=1}^n x_i}$$

This is the same as the binomial mass function, without the factor of

$$\binom{n}{\sum_{i=1}^n x_i}$$

This is because the bernoulli likelihood is the mass function of *any particular combination* of $x_1 \ldots x_n$, while the binomial mass function gives the probability of obtaining $\sum_{i=1}^n x_i$ 1's- which can happen $\binom{n}{\sum_{i=1}^n x_i}$ ways.

**Assignment 4, Question 4**: As on assignment 2, let $\epsilon_i \sim N(0, \sigma)$ be independent and identically distributed random variables. Let $\beta \in \mathbb{R}$ be a fixed, unknown constant and $x_i \in \mathbb{R}, i = 1 \ldots n$ be fixed, *known* quantities. Let $Y_i = \beta x_i + \epsilon_i, i = 1 \ldots n$.

(a) Find the MLE for $\beta$

(b) What is its *exact* sampling distribution? Don't use any limiting approximations; work it out exactly. You can answer this by remembering a question from assignment 1 that dealt with the distribution of a sum of independent normal random variables.

(c) Find the MLE for $\sigma^2$

(d) What is its *exact* sampling distribution? Don't use any limiting approximations; work it out exactly. You can answer this by remembering a question from assignment 1 that dealt with the distribution of a sum of squares of independent normal random variables.

(e) Evaluate the Observed Information and Fisher Information matrices for $\hat{\beta}, \hat{\sigma}^2$. Invert the Fisher Information matrix (using the formula for the inverse of a $2 \times 2$ matrix); do the resulting variance estimates agree with the ones you derived exactly? Why or why not?

*Solution*:

(a) Note that $Y_i \sim N(\beta x_i, \sigma^2)$. The likelihood is

$$L(\beta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \times \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^n (y_i - x_i\beta)^2}{\sigma^2}\right)$$

The log-likelihood is therefore

$$\ell(\beta) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - x_i\beta)^2$$

Differentiating with respect to $\beta$ yields

$$S(\beta) = \frac{1}{\sigma^2}\sum_{i=1}^n x_i(y_i - \beta x_i)$$

Setting to zero and solving gives

$$\hat{\beta} = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i^2}$$

(b) For its sampling distribution, note that $\hat{\beta} = \sum_{i=1}^n a_i y_i$, a linear combination of normals with

$$a_i = \frac{x_i}{\sum_{i=1}^n x_i^2}$$

So $\hat{\beta}$ is normal with

$$E(\hat{\beta}) = \sum_{i=1}^n a_i E(y_i) = \beta$$

$$Var(\hat{\beta}) = \sum_{i=1}^n a_i^2 Var(y_i) = \frac{1}{\sum_{i=1}^n x_i^2}\sigma^2$$

(c) The log-likelihood in the $\sigma^2$-dimension is

$$\ell(\sigma^2) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - x_i\beta)^2$$

The score function is

$$S(\sigma^2) = \frac{\partial \ell(\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (y_i - \beta x_i)^2$$

Setting to zero and solving gives

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \beta x_i)^2$$

Note that to use this, you would plug in $\hat{\beta}$ for $\beta$.

(d) For its sampling distribution, note that each $y_i - \beta x_i$ is a $N(0, \sigma)$ random variable. Assignment 1 had a question about the distribution of a sum of squared *standard* normal random variables. This is the $\chi^2$ distribution with $n$ degrees of freedom, although it doesn't matter whether you know the name or not; just stating the MGF or PDF is enough to answer the question:

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_n$$

This answer is for $\beta$ fixed; if you knew from another course the answer involving $\hat{\beta}$, that's fine too. The point of this question was to get you to identify the form of the estimator as being a sum of squared normals.

(e) The observed information is

$$\begin{aligned}
J_{\beta\beta}(\beta, \sigma^2) &= -\frac{\partial S(\beta)}{\partial \beta} = \frac{\sum_{i=1}^n x_i^2}{\sigma^2} \\
J_{\sigma^2\sigma^2}(\beta, \sigma^2) &= -\frac{\partial S(\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^n (y_i - \beta x_i)^2 \\
J_{\beta\sigma^2}(\beta, \sigma^2) &= -\frac{\partial S(\beta)}{\partial \sigma^2} = -\frac{\partial S(\sigma^2)}{\partial \beta} = \frac{1}{\sigma^4}\sum_{i=1}^n x_i (y_i - \beta x_i)
\end{aligned}$$

or in matrix form,

$$J(\beta, \sigma^2) = \begin{pmatrix} \frac{\sum_{i=1}^n x_i^2}{\sigma^2} & \frac{1}{\sigma^4}\sum_{i=1}^n x_i (y_i - \beta x_i) \\ \frac{1}{\sigma^4}\sum_{i=1}^n x_i (y_i - \beta x_i) & -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^n (y_i - \beta x_i)^2 \end{pmatrix}$$

(f) The Fisher Information is

$$I(\beta, \sigma^2) = E\left(J(\beta, \sigma^2)\right) = \begin{pmatrix} \frac{\sum_{i=1}^n x_i^2}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

The expectation is with respect to $y_i$, because $x_i$ are fixed and known. The expectation of a matrix is the matrix of expectations, so take the expectation of each term from $J$ and stack them in a matrix.

(g) Because the Fisher Information is diagonal, it is easy to invert:

$$I^{-1}(\beta, \sigma^2) = \begin{pmatrix} \frac{\sigma^2}{\sum_{i=1}^n x_i^2} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}$$

From this, we see that the asymptotic variance of the MLE $(\hat{\beta}, \hat{\sigma^2})$ is

$$\begin{aligned}
Var(\hat{\beta}) &= I^{-1}(\beta, \sigma^2)_{\beta\beta} = \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \\
Var(\hat{\sigma}^2) &= I^{-1}(\beta, \sigma^2)_{\sigma^2\sigma^2} = \frac{2\sigma^4}{n}
\end{aligned}$$

This agrees with what we found when we evaluated the exact sampling distributions for the estimators, because the data is Normally distributed. We spoke in class about how the asymptotic variance of the MLE for the normal distribution is actually exact.

**Assignment 4, Textbook question 7**

(a) We have $E(X) = \frac{1}{p}$. Plug in $\bar{X}$ for $E(X)$ and rearrange to get $\hat{p} = \frac{1}{\bar{X}}$.

(b) For all questions like this that ask for both the MLE and its asymptotic variance, you need to get

- Likelihood
- Log-likelihood
- Score function
- Observed information
- Fisher information

The likelihood is

$$L(p) = \prod_{i=1}^{n} P(X_i = x_i) = p^n (1-p)^{\sum_{i=1}^{n}(x_i - 1)}$$

The log-likelihood is

$$\ell(p) = \log L(p) = n \log p + \log(1-p) \left( \sum_{i=1}^{n} x_i - n \right)$$

The score function is

$$S(p) = \frac{\partial \ell}{\partial p} = \frac{n}{p} - \frac{\sum_{i=1}^{n} x_i - n}{1-p}$$

Setting to zero and solving for $p$ gives

$$\hat{p} = \frac{1}{\bar{X}}$$

(a) The observed information is

$$J(p) = -\frac{\partial S(p)}{\partial p} = \frac{n}{p^2} + \frac{n(\bar{X}-1)}{(1-p)^2}$$

The Fisher information is

$$I(p) = E(J(p)) = n \left( \frac{1}{p^2} + \frac{1}{p(1-p)} \right)$$

where we used the fact that $E(\bar{X}) = E(X) = \frac{1}{p}$. So the asymptotic variance of $\hat{p}$ is

$$Var(\hat{p}) \approx \frac{1}{I(\hat{p_0})} = \frac{p_0^2(1-p_0)}{n}$$

which you can get after some algebra to get rid of the rational functions in the denominator. It would also be correct to replace $p_0$ with a consistent estimator, i.e $\hat{p}$.