Homework 3: Mathematical Statistics (MATH-UA 234)

Due 10/06 at the beginning of class on Gradescope

Problem 1. *Let*

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad \vec{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix}$$

where $a_{i,j}$ are constants and $x_1, x_2, \ldots, x_n \sim N(0,1)$ are independent and identically distributed standard normal random variables.

(a) What is the covariance matrix $\vec{\Sigma}$ for $\vec{x}$? (Recall the (i,j)-entry of $\vec{\Sigma}$ is $\mathrm{Co}\mathbb{V}[x_i, x_j]$.

(b) Show that $\mathbb{E}[\vec{x}^\mathsf{T}\vec{A}\vec{x}] = \mathrm{tr}(\vec{A}) := a_{1,1} + a_{2,2} + \cdots + a_{n,n}$. (hint: write out the expression for $\vec{x}^\mathsf{T}\vec{A}\vec{x}$ as a sum over the entires of $\vec{A}$)

(c) Suppose $\vec{A}$ is diagonal; i.e. $a_{i,j} = 0$ for all $i \neq j$. Compute the variance of $\vec{x}^\mathsf{T}\vec{A}\vec{x}$. You may use the fact that x_i^2 is a Chi-square random variable with one degree of freedom so that $\mathbb{V}[x_i^2] = 2$.

This is an example of "stochastic trace estimation" which is an important algorithmic tool in a number of recent algorithms.

Solution.

(a) If $i \neq j$, x_i and x_j are independent so $[\vec{\Sigma}]_{i,j} = \mathrm{Co}\mathbb{V}[x_i, x_j] = 0$ for $i \neq j$. We also have $[\vec{\Sigma}]_{i,i} = \mathrm{Co}\mathbb{V}[x_i, x_i] = \mathbb{V}[x_i] = 1$. Thus, $\vec{\Sigma}$ is the identity matrix.

(b) Using the definition of matrix multiplication we find

$$\vec{x}^\mathsf{T}\vec{A}\vec{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}x_ix_j.$$

Now, using the linearity of the expectation,

$$\begin{aligned} \mathbb{E}[\vec{x}^\mathsf{T}\vec{A}\vec{x}] &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}\mathbb{E}[x_ix_j] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}\mathbb{E}[(x_i-0)(x_j-0)] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}\,\mathrm{Co}\mathbb{V}[x_i, x_j] \\ &= \sum_{i=1}^{n} a_{i,j}. \end{aligned}$$

problems with a textbook reference are based on, but not identical to, the given reference

(c) If $\vec{A}$ is diagonal,

$$\vec{x}^\top \vec{A}\vec{x} = \sum_{i=1}^{n} a_{i,i} x_i^2.$$

Since x_i are independent, using that $\mathbb{V}[x_i^2] = 2$,

$$\mathbb{V}[\vec{x}^\top \vec{A}\vec{x}] = \sum_{i=1}^{n} a_{i,i}^2 \mathbb{V}[x_i^2] = \sum_{i=1}^{n} a_{i,i}^2 2 = 2\sum_{i=1}^{n} a_{i,i}^2.$$

Problem 2 (Wasserstein 5.5). *Suppose $X_1, X_2, \dots, X_n \sim \mathrm{Ber}(p)$ are independent and identically distributed. Let $Z_n = (X_1^2 + X_2^2 + \cdots + X_n^2)/n$. Prove that*

(a) Z_n converges in probability to the constant random variable p.

(b) Z_n converges in quadratic mean to the constant random variable p.

Solution.

(a) Since convergence in quadratic mean implies convergence in probability, we simply need to prove (b).

Alternately, we could note that since the outputs of X_i are 0 and 1, $X_i^2 = X_i$. Thus, $Z_n = \bar{X}_n$ so the Law of large numbers implies $Z_n \to \mathbb{E}[X_i] = p$ in probability.

Finally, we could just consider the definition and use Chebyshev's inequality using that $\mathbb{V}[Z_n] = p(1-p)/n$ (as shown below). In particular, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}[|Z_n - p| > \epsilon] \leq \lim_{n\to\infty} \frac{\mathbb{V}[Z_n]}{\epsilon^2} = \lim_{n\to\infty} \frac{p(1-p)}{n\epsilon^2} = 0.$$

Note that this very close to a proof that convergence in quadratic mean implies convergence in probability.

(b) Since the X_i are independent, we can use a computation like done in lecture to get

$$\mathbb{V}[Z_n] = \mathbb{V}[\bar{X}_n] = \frac{\mathbb{V}[X_i]}{n} = \frac{p(1-p)}{n}.$$

Thus, since $\mathbb{E}[Z_n] = \mathbb{E}[\bar{X}_n] = p$,

$$\mathbb{E}[(Z_n - p)^2] = \mathbb{E}[(Z_n - \mathbb{E}[Z_n])^2] = \mathbb{V}[Z_n] = \frac{p(1-p)}{n} \to 0.$$

This implies convergence in quadratic mean.

Problem 3 (Wasserstein 6.1). *Suppose $X_1, X_2, \dots, X_n \sim \mathrm{Poisson}(\lambda)$ are independent and identically distributed and let $\hat{\lambda}_n = n^{-1}\sum_{i=1}^{n} X_i$ be our point estimator for λ. Find the bias $\mathrm{Bias}_n = \mathbb{E}[\hat{\lambda}_n] - \lambda$, standard error $se_n = \sqrt{\mathbb{V}[\hat{\lambda}_n]}$, and mean squared error $\mathrm{MSE}_n = \mathbb{E}[(\hat{\lambda}_n - \lambda)^2]$ of $\hat{\lambda}_n$.*

Solution. Here $\hat{\lambda}_n$ is simply the sample mean. For a Poisson random variable X with parameter λ, we know (or can easily compute) $\mathbb{E}[X] = \lambda$ and $\mathbb{V}[X] = \lambda$. Thus,

$$\mathrm{Bias}_n = \mathbb{E}[\hat{\lambda}_n] - \lambda = \lambda - \lambda = 0, \qquad \mathrm{se}_n = \sqrt{\mathbb{V}[\hat{\lambda}_n} = \sqrt{\frac{\lambda}{n}}, \qquad \mathrm{MSE}_n = \mathrm{Bias}_n^2 + \mathrm{se}_n^2 = \frac{\lambda}{n}.$$

Problem 4 (Wasserstein 6.2). *Suppose $X_1, X_2, \dots, X_n \sim \mathrm{Unif}(0, \theta)$ are independent and identically distributed and let $\hat{\theta}_n = \max\{X_1, X_2, \dots, X_n\}$ be our point estimator for θ.*

(a) *Write down the distribution for $\hat{\theta}_n$ (hint: we've already done a similar problem)*

(b) *Find the bias $\mathrm{Bias}_n = \mathbb{E}[\hat{\theta}_n] - \theta$, standard error $se_n = \sqrt{\mathbb{V}[\hat{\theta}_n]}$, and mean squared error $\mathrm{MSE}_n = \mathbb{E}[(\hat{\theta}_n - \theta)^2]$ of $\hat{\theta}_n$.*

Solution.

(a) This is similar to a problem on homework 1. In particular, we have that

$$\begin{aligned} F_{\hat{\theta}_n}(t) = \mathbb{P}[\hat{\theta}_n \le t] &= \mathbb{P}[\max\{X_1, \dots, X_n\} \le t] \\ &= \mathbb{P}[X_1 \le t, \dots, X_n \le t] \\ &= \mathbb{P}[X_1 \le t] \cdots \mathbb{P}[X_n \le t]. \\ &= F_{X_1}(t) \cdots F_{X_n}(t). \end{aligned}$$

Since the X_i are iid uniform random variables,

$$F_{X_i}(t) = \begin{cases} t/\theta & t \in [0, \theta] \\ 1 & t > \theta \\ 0 & t < 0 \end{cases}.$$

Thus,

$$F_{\hat{\theta}_n}(t) = \begin{cases} t^n/\theta^n & t \in [0, \theta] \\ 1 & t > \theta \\ 0 & t < 0 \end{cases}.$$

(b) We compute the density

$$f_{\hat{\theta}_n}(t) = F'_{\hat{\theta}_n}(t) = \begin{cases} nt^{n-1}/\theta^n & t \in [0, \theta] \\ 0 & t > \theta \\ 0 & t < 0 \end{cases}.$$

We can now directly compute

$$\mathbb{E}[\hat{\theta}_n] = \int t f_{\hat{\theta}_n}(t) \mathrm{d}t = \int_0^\theta t \frac{nt^{n-1}}{\theta^n} \mathrm{d}t = \frac{n}{\theta^n} \left[\frac{t^{n+1}}{n+1} \right]_{t=0}^{t=\theta} = \frac{n\theta}{n+1}.$$

Thus,

$$\mathrm{Bias}_n = \mathbb{E}[\hat{\theta}_n] - \theta = \frac{n\theta}{n+1} - \theta = \frac{-\theta}{n+1}.$$

To compute the variance, we first compute,

$$\mathbb{E}[\hat{\theta}_n] = \int t^2 f_{\hat{\theta}_n}(t) \mathrm{d}t = \int_0^\theta t^2 \frac{nt^{n-1}}{\theta^n} \mathrm{d}t = \frac{n}{\theta^n} \left[\frac{t^{n+2}}{n+2} \right]_{t=0}^{t=\theta} = \frac{n\theta^2}{n+2}.$$

We then have

$$\mathbb{V}[\hat{\theta}_n] = \mathbb{E}[\hat{\theta}_n^2] - \mathbb{E}[\hat{\theta}_n]^2 = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{n+1} = \frac{n\theta^2}{(n+1)^2(n+2)}.$$

Finally,

$$\text{se}_n = \sqrt{\mathbb{V}[\hat{\theta}_n]} = \sqrt{\frac{n\theta^2}{(n+1)^2(n+2)}}.$$

We can compute the MSE similar as above and get

$$\text{MSE}_n = \text{Bias}_n^2 + \text{se}_n^2 = \frac{\theta^2}{(n+1)^2} + \frac{n\theta^2}{(n+1)^2(n+2)} = \frac{n\theta + \theta^2(n+2)}{(n+1)^2(n+2)}.$$

Problem 5 (Wasserstein 6.3). *Suppose $X_1, X_2, \dots, X_n \sim \text{Unif}(0,\theta)$ are independent and identically distributed and let $\hat{\theta}_n = 2\overline{X}_n$ be our point estimator for θ. Find the bias $\text{Bias}_n = \mathbb{E}[\hat{\theta}_n] - \theta$, standard error $se_n = \sqrt{\mathbb{V}[\hat{\theta}_n]}$, and mean squared error $\text{MSE}_n = \mathbb{E}[(\hat{\theta}_n - \theta)^2]$ of $\hat{\theta}_n$.*

Solution. Here $\hat{\theta}_n$ is simply the sample mean *divided by two*. Thus, the expectation is half the expectation of the sample mean, and the variance is one quarter the variance of the sample mean. For a Uniform random variable X on $[0,\theta]$, we know (or can easily compute) $\mathbb{E}[X] = \theta/2$ and $\mathbb{V}[X] = \theta^2/12$. Thus,

$$\text{Bias}_n = \mathbb{E}[\hat{\lambda}_n] - \theta = 2(\theta/2) - \theta = 0, \qquad \text{se}_n = \sqrt{\mathbb{V}[\hat{\lambda}_n} = \sqrt{\frac{4\theta^2}{12n}} = \frac{\theta}{\sqrt{3n}}, \qquad \text{MSE}_n = \text{Bias}_n^2 + \text{se}_n^2 = \frac{\theta^2}{3n}.$$

Problem 6. *Suppose $X_1, X_2, \dots, X_n \sim F_{\mu,\sigma^2}$ are independent and identically distributed samples from some distribution F_{μ,σ^2} with mean μ and variance σ^2.*

Recall that

$$\hat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2, \qquad \hat{T}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

are both point estimators for the parameter σ^2.

(a) Compute $\mathbb{E}[\hat{S}_n^2]$ and $\mathbb{E}[\hat{T}_n^2]$.

(b) Which point estimator has smaller bias?

(c) Which point estimator has smaller standard error?

Solution.

(a) Define

$$Z = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

Using the linearity of expectation,

$$\mathbb{E}[Z] = \sum_{i=1}^{n}\mathbb{E}[(X_i - \bar{X}_n)^2] = \sum_{i=1}^{n}\mathbb{E}[(X_i-\mu)^2] - 2\mathbb{E}[(X_i-\mu)(\bar{X}_n-\mu)] + \mathbb{E}[(\bar{X}_n-\mu)^2].$$

Note that $\mathbb{E}[(X_i-\mu)^2] = \mathbb{V}[X_i] = \sigma^2$. Similarly, since $\mathbb{E}[\bar{X}_n] = \mu$, $\mathbb{E}[(\bar{X}_n-\mu)^2] = \mathbb{V}[\bar{X}_n] = \sigma^2/n$. To compute $\mathbb{E}[(X_i-\mu)(\bar{X}_n-\mu)]$, we can expand and use the linearity of expectation to get

$$
\begin{aligned}
\mathbb{E}[(X_i-\mu)(\bar{X}_n-\mu)] &= \mathbb{E}\left[(X_i-\mu)\sum_{i=1}^{n}\frac{1}{n}(X_j-\mu)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(X_i-\mu)(X_j-\mu)\right] \\
&= \frac{1}{n}\mathbb{E}[(X_i-\mu)^2] + \frac{1}{n}\sum_{j\neq i}\mathbb{E}[(X_i-\mu)(X_j-\mu)] \\
&= \frac{1}{n}\mathbb{V}[X_i] = \frac{\sigma^2}{n},
\end{aligned}
$$

where we have use that X_i and X_j are independent in the last line.

Putting this together, we find

$$
\mathbb{E}[(X_i-\mu)^2] = \sigma^2 - 2\frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \sigma^2 - \frac{\sigma^2}{n} = \sigma^2\frac{n-1}{n}
$$

so that

$$
\mathbb{E}\left[\sum_{i=1}^{n}(X_i-\bar{X}_n)^2\right] = \sum_{i=1}^{n}\sigma^2\frac{n-1}{n} = \sigma^2(n-1).
$$

Thus, we find

$$
\mathbb{E}[\hat{S}_n^2] = \mathbb{E}\left[\frac{Z}{n-1}\right] = \frac{1}{n-1}\sigma^2(n-1) = \sigma^2.
$$

and

$$
\mathbb{E}[\hat{T}_n^2] = \mathbb{E}\left[\frac{Z}{n}\right] = \frac{1}{n}\sigma^2(n-1) = \sigma^2\frac{n-1}{n}.
$$

(b) This problem was poorly worded, because it is unclear whether "smaller" meant smaller magnitude, or more negative. The intended question was about which was smaller in magnitude, which is the unbiased estimator $\hat{S}_n$.

(c) Note that

$$
\mathbb{V}[\hat{S}_n^2] = \mathbb{V}\left[\frac{Z}{n-1}\right] = \frac{1}{(n-1)^2}\mathbb{V}[Z], \qquad \mathbb{V}[\hat{T}_n^2] = \mathbb{V}\left[\frac{Z}{n}\right] = \frac{1}{n^2}\mathbb{V}[Z]
$$

Since $\mathbb{V}[Z] \geq 0$, we have $\mathbb{V}[\hat{S}_n^2] \geq \mathbb{V}[\hat{T}_n^2]$.

Note that computing $\mathbb{V}[Z]$ is actually really tedious: see for instance the proof on Mathworld.

Problem 7. *Describe of point estimation of a parameter which you noticed in a different part of* ***your life*** *(e.g. in other classes, on the subway, at the park, in the dorm, etc.).*