HW2, Theory of Inference 2016/7

Jonathan Rougier
School of Mathematics
University of Bristol UK

In the following questions, I show marks in square brackets, to give you an idea of the approximate tariff the question would carry in an exam.

Here are some questions about certification of algorithms for point-estimation. Note below that certification is always for an algorithm *and* a model. You do not need to hand in Q2d.

1. Consider the model in which $X_1, \dots, X_m$ are IID, which we express as

$$\mathcal{E}^m = \big\{\mathcal{X}^m, \Omega, f\big\}$$

where $f(x; \theta) = \prod_{i=1}^m f_1(x_i; \theta)$ for some f_1.

(a) Show that that if $m(\theta)$ is the expectation of X_1 (i.e. of each X_i), then

$$\tilde{\theta}_\pi(x) = \frac{1}{n} \sum_{j=1}^n x_{\pi_j} \tag{1}$$

is an unbiased estimator of $m(\theta)$ in $\mathcal{E}^m$, where $\pi = (\pi_1, \dots, \pi_n)$ is any size-n subset of $\{1, \dots, m\}$. Hint: take $\pi = (1, \dots, n)$, without loss of generality. [10 marks]

Answer. $\tilde{\theta}_\pi$ is an unbiased estimator of $m(\theta)$ exactly when $\mathrm{E}\{\tilde{\theta}_\pi(X); \theta\} = m(\theta)$ for all $\theta \in \Omega$. Take the hint, and let $\pi = (1, \dots, n)$. Then, for arbitrary

$\theta \in \Omega$,

$$\begin{aligned}
\mathrm{E}\{\tilde{\theta}_\pi(X);\theta\} &= \frac{1}{n}\sum_{j=1}^{n} \mathrm{E}(X_j;\theta) && \text{linearity of expectation} \\
&= \frac{1}{n}\sum_{j=1}^{n} m(\theta) && X\text{'s are IID} \\
&= m(\theta).
\end{aligned}$$

This holds for all θ, since the choice of θ was arbitrary. It holds for any π because in the IID model, $\mathrm{E}(X_i;\theta) = m(\theta)$ for all $i = 1, \dots, m$.

(b) Show that, for any fixed π,

$$s_\pi^2(x) = \frac{1}{n-1}\sum_{j=1}^{n} \left\{x_{\pi_j} - \tilde{\theta}_\pi(x)\right\}^2$$

is an unbiased estimator of the variance of X_1 for $\mathcal{E}^m$. [10 marks]

Answer. This is a classic result. Again, let $\pi = (1, \dots, n)$, so that

$$s_\pi^2(x) = \frac{1}{n-1}\sum\nolimits_j (x_j - \bar{x})^2,$$

where $\bar{x} = \tilde{\theta}(x) = (x_1 + \cdots + x_n)/n$. Now consider the 'obvious' estimator,

$$t^2(x) = \frac{1}{n}\sum\nolimits_j (x_j - \bar{x})^2 = \frac{1}{n}\sum\nolimits_j x_j^2 - \bar{x}^2,$$

after expanding out the square. We have

$$\mathrm{E}(X_j^2;\theta) = \mathrm{Var}(X_j;\theta) + \mathrm{E}(X_j;\theta)^2 = \sigma^2(\theta) + m(\theta)^2,$$

where $\sigma^2(\theta) :=$ $\mathrm{Var}(X_1;\theta)$; here we have used that the X's are IID. Likewise,

$$\mathrm{E}(\bar{X}^2;\theta) = \mathrm{Var}(\bar{X};\theta) - \mathrm{E}(\bar{X};\theta)^2 = \sigma^2(\theta)/n + m(\theta)^2,$$

using that the X's are IID and uncorrelated. Hence

$$\begin{aligned}
\mathrm{E}\{t^2(X);\theta\} &= \sigma^2(\theta) + m(\theta)^2 - \{\sigma^2(\theta)/n + m(\theta)^2\} \\
&= (1 - 1/n)\sigma^2(\theta) = \frac{n-1}{n}\sigma^2(\theta).
\end{aligned}$$

Hence, by the linearity of expectation, $\{n/(n-1)\}\, t^2(x) = s^2(x)$ is an unbiased estimator of $\sigma^2(\theta)$ under $\mathcal{E}^m$.

(c) Show that $\sqrt{s_\pi^2}$ is *not* an unbiased estimator of the standard deviation of X_1 for $\mathcal{E}^m$. Hint: Jensen's inequality. [10 marks]

Answer. Jensen's inequality states that if g is a convex function, then $\mathrm{E}\{g(X)\} \geq g(\mathrm{E}\{X\})$. A stronger version states that if g is strictly convex and $\mathrm{Var}(X) > 0$, then $\mathrm{E}\{g(X)\} > g(\mathrm{E}\{X\})$. $-\sqrt{(\cdot)}$ is a convex function (because $\sqrt{(\cdot)}$ is a concave function). Ignoring the trivial case, so that $\sigma^2(\theta) > 0$ for some $\theta \in \Omega$ (I am dropping the π subscript),

$$\mathrm{E}\{-\sqrt{s^2(X)}; \theta\} > -\sqrt{\mathrm{E}\{s^2(X); \theta\}} = -\sqrt{\sigma^2(\theta)},$$

for some $\theta \in \Omega$. Hence $\mathrm{E}\{\sqrt{s_\pi^2(X)}; \theta\} < \sqrt{\sigma^2(\theta)}$, showing that $\sqrt{s^2(x)}$ is not an unbiased estimator of $\sqrt{\sigma^2(\theta)}$ even though $s^2(x)$ *is* an unbiased estimator of $\sigma^2(\theta)$ for $\mathcal{E}^m$. This is the basic problem with unbiasedness as a certification: it does not extend to non-linear functions of unbiased estimators. Another problem is that unbiased estimators may not exist.

2. Consider the model $\mathcal{E} = \{\mathcal{X}, \Omega, f\}$. Point estimators of scalar functions of θ are often certified by their Mean Squared Error (MSE) function. If $\hat{g}$ is a point estimator of $g(\theta) \in \mathbb{R}$, then

$$\mathrm{MSE}_{\hat{g}}(\theta) := \mathrm{E}[\{\hat{g}(X) - g(\theta)\}^2; \theta]$$

is the MSE of $\hat{g}$.

(a) Express $\mathrm{MSE}_{\hat{g}}$ as an explicit equation involving the components of $\mathcal{E}$, and $\hat{g}$ and θ. [5 marks]

Answer. Working directly from the definition,

$$\mathrm{MSE}_{\hat{g}}(\theta) = \sum_{x \in \mathcal{X}} \big\{\hat{g}(x) - g(\theta)\big\}^2 \cdot f(x; \theta).$$

(b) Show that

$$\mathrm{MSE}_{\hat{g}}(\theta) = \mathrm{Var}_{\hat{g}}(\theta) + \mathrm{bias}_{\hat{g}}(\theta)^2,$$

giving exact definitions for the two functions on the righthand side. Hint: consider introducing the quantity $\bar{g}(\theta) = \mathrm{E}\{\hat{g}(X); \theta\}$. [10 marks]

Answer. Taking the hint,

$$\begin{aligned}\mathrm{MSE}_{\hat{g}}(\theta) &= \mathrm{E}[\{\hat{g}(X) - \bar{g}(\theta) + \bar{g}(\theta) - g(\theta)\}^2; \theta] \\ &= \mathrm{E}[\{\hat{g}(X) - \bar{g}(\theta)\}^2; \theta] + \mathrm{E}[\{\bar{g}(\theta) - g(\theta)\}^2; \theta] + \text{cross product term} \\ &= \mathrm{Var}_{\hat{g}}(\theta) + \mathrm{bias}_{\hat{g}}(\theta)^2 + \text{cross product term}\end{aligned}$$

according to the usual definitions. So we have to show that the cross product term is zero for all $\theta \in \Omega$:

$$\begin{aligned}\mathrm{E}[\{\hat{g}(X) - \bar{g}(\theta)\}\{\bar{g}(\theta) - g(\theta)\}; \theta] &= \{\bar{g}(\theta) - g(\theta)\} \times \mathrm{E}[\{\hat{g}(X) - \bar{g}(\theta); \theta\} \\ &= \{\bar{g}(\theta) - g(\theta)\} \times [\mathrm{E}\{\hat{g}(X); \theta\} - \bar{g}(\theta)] \\ &= \{\bar{g}(\theta) - g(\theta)\} \times 0 = 0.\end{aligned}$$

(c) Compute the MSE of the estimator $\tilde{\theta}$ in (1). [10 marks]

Answer. The bias is zero, as we have already shown, so

$$\begin{aligned}\mathrm{MSE}_{\tilde{\theta}}(\theta) &= \mathrm{Var}_{\tilde{\theta}}(\theta) \\ &= \mathrm{E}[\{\tilde{\theta}(X) - \theta\}^2; \theta] \\ &= \mathrm{Var}\{\tilde{\theta}(X); \theta\} \\ &= \sigma^2(\theta)/n,\end{aligned}$$

where $\sigma^2(\theta) = \mathrm{Var}(X_1; \theta)$, as above.

(d) (Not examinable.) Let $\mathcal{E}^m$ be the IID model. An estimator $\hat{g}_m$ of $g(\theta)$ is certified as *consistent* for $\mathcal{E}^m$ exactly when

$$\forall \theta \in \Omega, \forall \epsilon > 0 \quad \lim_{m \to \infty} \Pr\left\{|\hat{g}_m(X_{1:m}) - g(\theta)| \geq \epsilon; \theta\right\} = 0.$$

Show that $\hat{g}_m$ is consistent if and only if, for all $\theta \in \Omega$, $\hat{g}_m$ converges to a point in Ω and its bias converges to zero.

Answer.

This result is not provable without a clear definition of 'converge to a point'. Here is what we will assume (the 'if' branch):

(i) Asymptotic unbiased:

$$\forall \theta : \lim_{m \to \infty} |\mathrm{E}\{\hat{g}_m(X_{1:n}); \theta\} - g(\theta)| = 0.$$

(ii) Converge to a point:

$$\forall\theta, \exists g_\infty, \forall\epsilon > 0 : \lim_{m\to\infty} \Pr\{|\hat{g}_m(X_{1:m}) - g_\infty| \geq \epsilon; \theta\} = 0.$$

(ii) implies that $\lim_{m\to\infty} \mathrm{E}\{\hat{g}_m(X_{1:m}); \theta\} = g_\infty$, because if a distribution collapses to a point, then that point must be the expectation. (i) then implies that $g_\infty = g(\theta)$. Substituting back into (ii) then gives

$$\forall\theta, \forall\epsilon > 0 : \lim_{m\to\infty} \Pr\{|\hat{g}_m(X_{1:m}) - g(\theta)| \geq \epsilon; \theta\} = 0,$$

as required.

I don't think the converse is true. Apparently, you cannot believe everything you read on Wikipedia.