Mathematical Statistics Solutions: HW 4

October 19, 2022

Exercise 1

Part (a)

If $\rho(t) = 2e^{-t^2/2}$, then we want to prove $\mathbb{P}_\theta(|\hat{\theta} - \theta| \geq t) \leq \rho(t\sqrt{n}) = 2e^{-nt^2/2}$. In the Bernoulli case, this follows from an application of Hoeffding's inequality, since the random variables lie in $[0, 1]$. The Gaussian case follows from the Chernoff-bound analogue.

Part (b)

Fix $\varepsilon > 0$, with $\mathbb{P}_\theta(|\hat{\theta} - \theta| \geq t) \leq \rho(t\sqrt{n})$. Since $\lim_{n\to\infty} \rho(t\sqrt{n}) = 0$, it follows that

$$0 \leq \lim_{n\to\infty} \mathbb{P}_\theta(|\hat{\theta} - \theta| \geq t) \leq \lim_{n\to\infty} \rho(t\sqrt{n}) = 0\,.$$

Part (c)

$$\begin{aligned}
\mathbb{P}_\theta(\theta \in \hat{C}) &= 1 - \mathbb{P}_\theta(\theta \notin \hat{C}) \\
&= 1 - \mathbb{P}_\theta(|\hat{\theta} - \theta| \leq \rho^{-1}(\alpha)/\sqrt{n} \\
&\geq 1 - \rho(\sqrt{n}\rho^{-1}(\alpha)/\sqrt{n}) \\
&= 1 - \rho(\rho^{-1}(\alpha)) = 1 - \alpha\,.
\end{aligned}$$

For a fixed θ, the probability that it lies in $\hat{C}$ is at least $1 - \alpha$ (many other verbose ways to write this).

Part (d)

For all $\theta \in \Theta_1$,

$$\begin{aligned}
\mathbb{P}_\theta(\psi = 1) &= \mathbb{P}_\theta(\exists \theta_1 \in \Theta_1 \text{ and } |\hat{\theta} - \theta_1| \leq \delta) \\
&= 1 - \mathbb{P}_\theta(\forall \theta_1 \in \Theta_1,\ |\hat{\theta} - \theta_1| \geq \rho^{-1}(\alpha)/\sqrt{n}) \\
&\geq 1 - \alpha,
\end{aligned}$$

by **(c)**. Similarly, for all $\theta \in \Theta_0$,

$$\begin{aligned}
\mathbb{P}_\theta(\psi = 0) &= \mathbb{P}_\theta(\forall \theta_1 \in \Theta_1,\ |\hat{\theta} - \theta_1| \geq \delta) \\
&\geq \mathbb{P}_\theta(|\hat{\theta} - \theta_0| < \delta) \\
&\geq \mathbb{P}_\theta(|\hat{\theta} - \theta_0| < \rho^{-1}(\alpha)/\sqrt{n}) \\
&\geq 1 - \alpha
\end{aligned}$$

Part (e)

Let $\Theta_0 := [0, \frac{1}{4}]$ and $\Theta_1 := [\frac{3}{4}, 1]$, and define the Bernoulli family over $\Theta := \Theta_0 \cup \Theta_1$, which satisfies equation (4.4), and suppose the true parameter of the data-generating process is $\theta = 3/4$ (generating $X_1, \ldots, X_n$ observations) and let $\hat{\theta} = \bar{X}_n$.

Now compute the following probability,

$$\mathbb{P}_\theta(\tilde{\psi} = 1) = \mathbb{P}_\theta(\hat{\theta} \in [3/4, 1]) = \mathbb{P}_\theta(\hat{\theta} \geq \theta) \leq 2/3,$$

for n sufficiently large (by CLT calculations). It is therefore impossible to guarantee $\mathbb{P}_\theta(\tilde{\psi} = 1) \geq 1 - \alpha$ for $\alpha < 1/3$.

Exercise 2

$\sigma(\cdot)$ is a continuous function so if $T_n \xrightarrow{p} \theta$ then by continuous mapping theorem, $1/\sigma(T_n) \xrightarrow{p} 1/\sigma(\theta)$. We also have that $\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta)$ so by Slutsky's theorem,

$$\frac{\sqrt{n}(T_n - \theta)}{\sigma(T_n)} \xrightarrow{d} N(0, \sigma^2(\theta)/\sigma^2(\theta) = N(0, 1).$$

Thus,

$$\begin{aligned}
\mathbb{P}_\theta(\theta \in \hat{C}_n) &= \mathbb{P}_\theta(\theta \in [T_n - \sigma(\theta) z_{\alpha/2}/\sqrt{n}, T_n + \sigma(\theta) z_{\alpha/2}/\sqrt{n}]) \\
&= \mathbb{P}_\theta\left(\left|\frac{\sqrt{n}(T_n - \theta)}{\sigma(T_n)}\right| \leq z_{\alpha/2}\right) \\
&\to \mathbb{P}_\theta(|Z| \leq z_{\alpha/2}) \quad \text{(in the limit } n \to \infty) \\
&= 1 - \alpha,
\end{aligned}$$

which implies that $\hat{C}_n$ is an asymptotic confidence interval.

Exercise 3

Part (a)

Recall that the DKW inequality gives,

$$\mathbb{P}\left(\sup_{t \in \mathbb{R}} \left|F(t) - \hat{F}_n(t)\right| \geq s\right) \leq 2e^{-2ns^2}.$$

Set $\alpha = 2e^{-2ns^2}$, we get

$$\mathbb{P}\left(\sup_{t \in \mathbb{R}} \left|F(t) - \hat{F}_n(t)\right| \geq \sqrt{\frac{\log(2/\alpha)}{2n}}\right) \leq \alpha.$$

Therefore, for any $t \in \mathbb{R}$, $(\underline{F}(t), \overline{F}(t))$ gives a $1 - \alpha$ confidence band for F, where

$$\underline{F}(t) = \hat{F}_n(t) - \sqrt{\frac{\log(2/\alpha)}{2n}}, \quad \overline{F}(t) = \hat{F}_n(t) + \sqrt{\frac{\log(2/\alpha)}{2n}}.$$

Part (b)

For any $t \in \mathbb{R}$, if $\underline{F}(t) < 0$ we can define the following sets,

$$C := (\underline{F}(t), \overline{F}(t)), \;\; C_1 := (\underline{F}(t), 0), \; C_2 := [0, \overline{F}(t)).$$

We have $C = C_1 \cup C_2$, and since $F(t) \geq 0$, we have $\mathbb{P}(F(t) \in C_1) = 0$. Therefore, $\mathbb{P}(F(t) \in C_2) = \mathbb{P}(F(t) \in C) \geq 1 - \alpha$. Thus $(\max\{\underline{F}, 0\}, \overline{F})$ is a $1 - \alpha$ confidence band for F. Using the same argument for the case $\overline{F}(t) \geq 1$, we conclude that $(\max\{\underline{F}, 0\}, \min\{\overline{F}, 1\})$ is a $1 - \alpha$ confidence band for F.

Exercise 4

Part (a)

$S_n := \frac{1}{n}\sum_{i=1}^n X_i \sim N(\mu, \frac{1}{n}\sigma^2)$, so $(S_n - \mu)/\sigma \sim N(0,1)$, which is pivotal.

Part (b)

For all $\theta \in \Theta$,

$$\begin{aligned} P_\theta[\theta \in C(X_1, \ldots, X_n)] &= P_\theta[\underline{c} \leq g(X_1, \ldots, X_n, \theta) \leq \overline{c}] \\ &\geq 1 - \alpha \end{aligned}$$

where we have used the given property of $g(X_1, \ldots, X_n, \theta)$, and that it holds for any θ since g is pivotal.

Part (c)

We can compute these values because $g(\omega, \theta)$ is independent of θ. For example, in the case where the pivotal quantity is the standard normal Gaussian (like in **(4a)**), one can pre-compute $z_{\alpha/2}$ such that $\mathbb{P}(|Z| \geq z_{\alpha/2}) = \alpha$, for $Z \sim N(0,1)$.