Math Stats

Instructor: Jonathan Niles-Weed

Homework 4
Due: Sunday October 9, 11:59pm via NYU Gradescope

Collaborated with Jonah Potzcobutt, and Andre Chen.

1. This problem shows how concentration bounds can be used to obtain estimators, confidence sets, and tests. Suppose we observe n i.i.d. samples from a parametric model $\mathcal{P} = \{\mathbb{P}\{\theta\} : \theta \in \Theta \subseteq \mathbb{R}\}$, and assume the existence of an estimator $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ such that:

$$\mathbb{P}_\theta \left\{ |\hat{\theta} - \theta| \geq t \right\} \leq \rho(\sqrt{n}t) \quad \forall t \leq 0, \theta \in \Theta$$

Where $\rho : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a strictly decreasing, continuous function satisfying $\lim_{s\to\infty} \rho(s) = 0$. Note that the meaning of (4.4) is as follows: our data $\omega = (X_1, \ldots, X_n)$ lives in the sample space Ω. No matter which probability measure $\mathbb{P}_\theta \in \mathcal{P}$ the space is equipped with, the random variable $\hat{\theta}(\omega)$ under that measure satisfies (4.4).

(a) Show that under the Bernoulli model, with $X_1, \ldots, X_n \sim Bern(\theta)$, and the Gaussian model, with $X_1, \ldots X_n \sim \mathcal{N}(\theta, 1)$, the estimator $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n X_i$ satisfies (4.4) with $\rho(t) = 2e^{-t^2/2}$.

What we want to show is:

$$\mathbb{P}_\theta \left\{ |\hat{\theta} - \theta| \geq t \right\} \leq \rho(\sqrt{n}t) = 2e^{-t^2 n/2} \quad \forall t \leq 0, \theta \in \Theta$$

Using the Chernoff bound for sums with i.i.d. random variables we have:

$$\mathbb{P} \left\{ (|\hat{\theta} - \theta|) \geq t \right\} \leq 2e^{\frac{-t^2 n}{2\sigma^2}} \quad \forall \lambda \geq 0$$

Since our random variable is bounded $\theta \in [0, 1]$, we have shown in homework that its $\frac{b-a}{4}$-subgaussian, and using a even more relaxed bound, its also $(b - a)$-subgaussian. Using the weaker bound we have: $\sigma^2 \leq (b - a)/4 \leq (b - a) = 1 - 0 = 1$, and we can use the bound in our definition definition of subgaussianity, substituting $\sigma^2 = 1$ as our variance.

Therefore, we have shown that for a Bernoulli the following expression holds:

$$\mathbb{P}_\theta \left\{ |\hat{\theta} - \theta| \geq t \right\} \leq \rho(\sqrt{n}t) = 2e^{-\frac{t^2 n}{2}} \quad \forall t \geq 0, \theta \in \Theta$$

Thus showing 4.4

We can make a similar argument for a Gaussian centered $\mathcal{N}(\theta, 1)$ using the definition of subgaussianty, and the fact that a Guassian is σ^2-subgaussian. In our case, since the variance of our gaussian is 1, then we have 1-subgaussian quantity. And thus, our inequality holds for this gaussian as well.

$$\mathbb{P}_\theta \left\{ |\hat{\theta} - \theta| \geq t \right\} \leq \rho(\sqrt{n}t) = 2e^{-\frac{t^2 n}{2}} \quad \forall t \geq 0, \theta \in \Theta$$

(b) Show that (4.4) implies $\hat{\theta} \xrightarrow{p} \theta$ as $n \to \infty$. This property is known as consistency.

As $n \to \infty$ then the quantity of $2e^{-\infty} \to 0$. Since the RHS goes to 0, then the left, which is less than or equal to the right, must also go to 0. Inspecting the LHS we see that since $t \geq 0$, we must have that for the LHS to be 0 $|\hat{\theta} - \theta| \to 0$, meaining $\hat{\theta} \xrightarrow{p} \theta$. We can also see this is apparant from our definition of ρ, as $\rho(s)$ is a decreasing function, and as $s \to \infty$ then $\rho(s) \to 0$.

(c) Fix $\alpha \in (0, 1)$. Show that if we define:

$$\hat{C} := \left[\hat{\theta} - \rho^{-1}(\alpha)/\sqrt{n}, \hat{\theta} + \rho^{-1}(\alpha)/\sqrt{n} \right]$$

Then $\mathbb{P}_\theta \left\{ \theta \in \hat{C} \right\} \geq 1 - \alpha$. State carefully the interpration of this statement. Such a set is known as a $1 - \alpha$ confidence interval.

We first inspect the quantity that $\mathbb{P}_\theta \left\{ \theta \notin \hat{C} \right\}$. For this to be true, we would need that $|\hat{\theta} - \theta| \geq \rho^{-1}(\alpha)/\sqrt{n}$. Therefore we can express:

$$\mathbb{P}_\theta \left\{ \theta \notin \hat{C} \right\} = \mathbb{P} \left\{ |\hat{\theta} - \theta| \geq \rho^{-1}(\alpha)/\sqrt{n} \right\} \quad \mathbb{P}_\theta \left\{ \theta \in \hat{C} \right\} = 1 - \mathbb{P}_\theta \left\{ \theta \notin \hat{C} \right\}$$

Using our previous bound we have that:

$$\mathbb{P}_\theta \left\{ \theta \notin \hat{C} \right\} = \mathbb{P} \left\{ |\hat{\theta} - \theta| \geq t \right\} \leq \rho(\sqrt{n}t)$$

Plugging in $t = \rho^{-1}(\alpha)\sqrt{n}$ then the right hand side simplifies to α, therefore we have:

$$\mathbb{P}_\theta \left\{ \theta \notin \hat{C} \right\} = \mathbb{P} \left\{ |\hat{\theta} - \theta| \geq \rho^{-1}(\alpha)\sqrt{n} \right\} \leq \alpha$$

Therefore,

$$\mathbb{P}_\theta \left\{ \theta \in \hat{C} \right\} = 1 - \mathbb{P}_\theta \left\{ \theta \notin \hat{C} \right\} \geq 1 - \alpha$$

We can interpret this statement that the probability that confidence interval $\hat{C}$ constructed around $\hat{\theta}$ contains the true parameter θ that we are estimating with n i.i.d. samples from $\mathcal{P} = P_\theta$ at least $1 - \alpha$.

(d) Let $\Theta_0, \Theta \subseteq \Theta$ be separate, in the sense that:

$$|\theta_0 - \theta_1| \geq 2\delta > 0 \quad \forall \theta_0 \in \Theta_0, \theta_1 \in \Theta_1.$$

Consider the test:

$$\psi = \mathbb{1}_{\exists \theta_1 \in \Theta_1 s.t. |\hat{\theta} - \theta_1 \leq \delta|}$$

Show that if $\delta > \rho^{-1}(\alpha)/\sqrt{n}$, then:

$$\mathbb{P}_\theta\{\psi = i\} \geq 1 - \alpha \quad \forall \theta \in \Theta_i, i \in \{0, 1\}$$

We can start by taking probability that $\psi = 1$

$$\begin{aligned} \mathbb{P}_\theta(\psi = 1) &= \mathbb{P}\left\{|\hat{\theta} - \theta_1| \leq \delta\right\} \\ &= \mathbb{P}\left\{|\hat{\theta} - \theta_1| \leq \rho^{-1}(\alpha)/\sqrt{n}\right\} \geq 1 - \alpha \quad \textit{part c)} \end{aligned} \tag{1}$$

We know that θ_0, θ_1 are disjoint and do not necessarily exhaust Θ as $\theta_0, \theta_1 \subseteq \Theta$ we can show:

$$\begin{aligned} \mathbb{P}\left\{|\hat{\theta} - \theta_0| \leq \delta\right\} + \mathbb{P}\left\{|\hat{\theta} - \theta_1| \leq \delta\right\} &\leq 1 \\ \mathbb{P}\left\{|\hat{\theta} - \theta_0| \leq \delta\right\} &\leq 1 - \mathbb{P}\left\{|\hat{\theta} - \theta_1| \leq \delta\right\} = \mathbb{P}_\theta(\psi = 0) \end{aligned} \tag{2}$$

Taking this:

$$\begin{aligned} \mathbb{P}_\theta(\psi = 0) &\geq 1 - \mathbb{P}\left\{|\hat{\theta} - \theta_0| \geq \delta\right\} \\ &\geq 1 - \mathbb{P}\left\{|\hat{\theta} - \theta_0| \geq \rho^{-1}(\alpha)/\sqrt{n}\right\} \\ &\geq 1 - \alpha \end{aligned} \tag{3}$$

The last step references what we found in part c) where:

$$\mathbb{P}\left\{|\hat{\theta} - \theta| \geq \rho^{-1}(\alpha)/\sqrt{n}\right\} \leq \alpha \quad \textit{and} \quad 1 - \mathbb{P}\left\{|\hat{\theta} - \theta| \geq \rho^{-1}(\alpha)/\sqrt{n}\right\} \geq 1 - \alpha$$

(e) Would the superficially similar test:

$$\tilde{\psi} = \mathbb{1}_{\hat{\theta} \in \Theta_1}$$

Yield the same guarantee? Why or why not?

This would not guarantee as the test does nothing other than show that $\hat{\theta} \in \Theta$, and does not convey if $\hat{\theta}$ is within $\rho^{-1}(\alpha)/\sqrt{n}$ of θ_1. Therefore, we could not make the same guarantees that we do in problem c and d, as we cannot apply the bounds: $\mathbb{P}\left\{|\hat{\theta} - \theta_1| \leq \rho^{-1}(\alpha)/\sqrt{n}\right\}$ which actually yield useful quantities.

This exercise justifies the attention we paid to concentration inequalities in the first lecture: with good concentration inequalities, we can estimate, create confidence sets, and test.

2. Suppose that we observe n i.i.d. samples from a parametric model $(\mathbb{R}, \mathcal{P})$, and suppose that under any $\mathbb{P}_\theta \in \mathcal{P}$, a sequence of statistics Tn satisfies a CLT centered at the parameter θ, i.e.,

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta)) \quad \forall \theta \in \Theta$$

for some function $\sigma : \mathbb{R} \to \mathbb{R}_{>0}$. Fix a $\alpha \in (0, 1)$ A sequence of sets $\hat{C}_n$ is called an asymptotic $(1 - \alpha)$ confidence set if:

$$\mathbb{P}_\theta \left\{ \theta \in \hat{C}_n \right\} \xrightarrow[n \to \infty]{} 1 - \alpha \quad \forall \theta \in \Theta$$

Define $z_{\alpha/2}$ to be the unique positive real number satisfying:

$$\mathbb{P}\left\{ |Z| \geq z_{\alpha/2} \right\} = \alpha \quad Z \sim \mathcal{N}(0, 1)$$

Assuming the function σ is continuous, show that the test:

$$\hat{C} := \left[T_n - \sigma(T_n) z_\alpha/2/\sqrt{n}, [T_n + \sigma(T_n) z_\alpha/2/\sqrt{n} \right]$$

is asymptotic $(1 - \alpha)$ confidence interval. (Hint: use Slutzky's theorem and the continuous mapping theorem.)

Answer:

We can pick a $g(x, y)$ such that:

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d} \mathcal{N}(0, g(x, y)'^2 \sigma^2(\theta)) = \mathcal{N}(0, 1) \quad \forall \theta \in \Theta$$

Choosing $g\left(x, \sigma(\theta)\right) = \frac{x}{\sigma(\theta)}$ will do nicely as $g(x)' = \frac{1}{\sigma(x)}$ then:

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d} \mathcal{N}(0, 1) \quad \forall \theta \in \Theta$$

By definition, we then have:

$$\mathbb{P}\left\{ |Z| \geq z_{\alpha/2} \right\} = \alpha \quad and \quad \mathbb{P}\left\{ |Z| \leq z_{\alpha/2} \right\} = 1 - \alpha \quad Z \sim \mathcal{N}(0, 1)$$

As we are talking about asymptotics, we can replace $|Z|$ with $|\sqrt{n}(g(T_n) - g(\theta)|$ when discussing probabilities as $n \to \infty$:

$$
\begin{aligned}
\mathbb{P}\left\{|Z| \le z_{\alpha/2}\right\} &= 1-\alpha \\
\mathbb{P}\left\{\left|\frac{\sqrt{n}}{\sigma(\theta)}(T_n-\theta)\right| \le z_{\alpha/2}\right\} &\xrightarrow[n\to\infty]{} 1-\alpha \\
\mathbb{P}\left\{\frac{\sqrt{n}}{\sigma(\theta)}(T_n-\theta) \le z_{\alpha/2}\right\} + \mathbb{P}\left\{\frac{\sqrt{n}}{\sigma(\theta)}(T_n-\theta) \ge -z_{\alpha/2}\right\} &\xrightarrow[n\to\infty]{} 1-\alpha \\
\mathbb{P}\left\{T_n - \sigma(\theta)z_{\alpha/2}/\sqrt{n} \le \theta\right\} + \mathbb{P}\left\{T_n + \sigma(\theta)z_{\alpha/2}/\sqrt{n} \ge \theta\right\} &\xrightarrow[n\to\infty]{} 1-\alpha \\
\mathbb{P}\left\{T_n - \sigma(\theta)z_{\alpha/2}/\sqrt{n} \le \theta \le T_n + \sigma(\theta)z_{\alpha/2}/\sqrt{n}\right\} &\xrightarrow[n\to\infty]{} 1-\alpha \\
\mathbb{P}\left\{T_n - \sigma(T_n)z_{\alpha/2}/\sqrt{n} \le \theta \le T_n + \sigma(T_n)z_{\alpha/2}/\sqrt{n}\right\} &\xrightarrow[n\to\infty]{} 1-\alpha \\
\mathbb{P}_\theta\left\{\theta \in \hat{C}_n\right\} \xrightarrow[n\to\infty]{} 1-\alpha \quad \forall\theta\in\Theta \quad \square
\end{aligned}
\tag{4}
$$

Where we can swap in $\sigma(T_n)$ for $\sigma(\theta)$ due to the continuous mapping theorem where $T_n \to \theta$ as $n \to \infty$, and therefore $\sigma(T_n) \to \sigma(\theta)$ as $n \to \infty$.

3. Consider n i.i.d. samples from the fully nonparametric model ($\mathbb{R}$, all probability distributions on $\mathbb{R}$). A pair of functions $\underline{F}, \overline{F} : \mathbb{R} \to \mathbb{R}$ constructed from the data is a $1-\alpha$ confidence band for the CDF F if:

$$
\mathbb{P}_F\left\{\underline{F}(t) \le F(t) \le \overline{F}(t)\ \forall t \in \mathbb{R}\right\} \ge 1-\alpha \quad \forall \mathbb{P}_F \in \mathcal{P},
$$

where $\mathbb{P}_F$ represents the probability measure with CDF F.

(a) Use (2.6) to construct a $1-\alpha$ confidence band for F.

From (2.6) we have:

$$
\mathbb{P}\left\{\sup_{t\in\mathbb{R}} |F(t) - \hat{F}_n(t)| \ge s\right\} \le 2e^{-2ns^2} \quad \forall s \ge 0
$$

Setting $s = \sqrt{\frac{log(2/\alpha)}{2n}}$ Then we have:

$$
\mathbb{P}\left\{\sup_{t\in\mathbb{R}} |F(t) - \hat{F}_n(t)| \ge s\right\} \le \alpha \quad \forall s \ge 0
$$

And

$$
\mathbb{P}\left\{\sup_{t\in\mathbb{R}} |F(t) - \hat{F}_n(t)| \le s\right\} \ge 1-\alpha \quad \forall s \ge 0
$$

Follows.

Let $\underline{F}(t) = F(t) - s$ and $\overline{F}(t) = F(t) + s$, we can reexpress $\mathbb{P}_F\left\{\underline{F}(t) \le F(t) \le \overline{F}(t)\ \forall t \in \mathbb{R}\right\}$ as:

$$
\begin{aligned}
\mathbb{P}_F\left\{\underline{F}(t) \le F(t) \le \overline{F}(t)\ \forall t \in \mathbb{R}\right\} &= \mathbb{P}\left\{\inf_{t\in\mathbb{R}} \hat{F}_n(t) - s \le F(t)\right\} + \mathbb{P}\left\{\sup_{t\in\mathbb{R}} \hat{F}_n(t) + s \ge F(t)\right\} \\
&= \mathbb{P}\left\{\sup_{t\in\mathbb{R}} |\hat{F}_n(t) - F(t)| \le s\right\} \ge 1-\alpha
\end{aligned}
\tag{5}
$$

We have constructed a $1-\alpha$ confidence band for F using:

$$\mathbb{P}_F\left\{\underline{F}(t) \leq F(t) \leq \overline{F}(t)\ \forall t \in \mathbb{R}\right\} = 1 - \mathbb{P}\left\{\sup_{t\in\mathbb{R}} |\hat{F}_n(t) - F(t)| \geq s\right\} \geq 1-\alpha$$

(b) Show that if $(\underline{F}, \overline{F})$ is a $1-\alpha$ confidence band for F, then $(\max\{\underline{F}, 0\}, \min\{\overline{F}, 1\})$ is as well. Therefore, the confidence band constructed in part (a) can always be truncated (if necessary) so that both $0 \leq \underline{F}(t) \leq \overline{F}(t) \leq 1$ for all $t \in \mathbb{R}$

This follows from the definition of a CDF being bound between [0,1]. Therefore, $\underline{F}$ must be greater than or equal to 0, and $\overline{F}$ must be less than or equal to 1. As $t \to -\infty$ then $\underline{F}(t) \to 0$ and conversely $t \to \infty$ then $\overline{F}(t) \to 1$. In short, this is trivial as $F(t) \in [0,1] \quad \forall t$

4. Given a statistical model $(\Omega, \mathcal{P})$, a function $g : \Omega \times \theta \to \mathbb{R}$ is called pivotal if:

$$\mathbb{P}_\theta\{g(\omega,\theta) \leq t\} = \mathbb{P}_{\theta'}\{g(\omega,\theta') \leq t\} \quad \forall t \in \mathbb{R}, \theta, \theta' \in \Theta,$$

that is, if the distribution of the random variable $g(\omega,\theta)$ under $\mathbb{P}_\theta$ does not depend on θ. Note that $g(\omega,\theta)$ is not a statistic, because it is not a function of the data alone.

(a) Consider the Gaussian model where $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$. Show that $(\frac{1}{n}\sum_{i=1}^n X_i - \mu)/\sigma$ is pivotal.

We know from CLT that:

$$\sqrt{n}(T_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \quad and \quad (T_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2/n)$$

Where $T_n = \frac{1}{n}\sum_{i=1}^n X_i$. Picking $g(\omega,\theta) = \frac{\frac{1}{n}\sum_{i=1}^n X_i - \mu}{\sigma}$:

$$g(\omega,\theta) = \frac{(\frac{1}{n}\sum_{i=1}^n X_i - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1/n)$$

Where $X_i, \ldots, X_n \sim \omega$ and $\mu, \sigma \in \theta$. We see that the results is a distribution that does not depend on μ or σ. As θ does not impact the resulting distribution, in fact, each resulting distribution is the same:

$$\mathcal{N}(0, 1/n) = \mathbb{P}_\theta\{g(\omega,\theta) \leq t\} = \mathbb{P}_{\theta'}\{g(\omega,\theta') \leq t\} = \mathcal{N}(0, 1/n)$$

Since the results have the same CDF, $g(\omega,\theta)$ is a pivotal function.

(b) Suppose that $\underline{c}, \overline{c} \in \mathbb{R}$ satisfy:

$$\mathbb{P}_\theta\{\underline{c} \leq g(\omega,\theta) \leq \overline{c}\} \leq 1-\alpha$$

Show that the set

$$C := \{\theta \in \Theta : g(\omega, \theta) \in [\underline{c}, \overline{c}]\}$$

is a $1 - \alpha$ confidence set.

(Trivial) We just showed that a pivotal $g(\omega, \theta)$ yields the same cdf for all $\theta \in \Theta$. Therefore, independent of Theta, $C := \{\theta \in \Theta : g(\omega, \theta) \in [\underline{c}, \overline{c}]\}$ holds.

(c) Why can $\underline{c}, \overline{c}$ satisfying (4.5) without knowledge of θ?

As $g(\omega, \theta)$ is pivotal, the resulting CDF is the same for all $\theta \in \Theta$, so in a sense the distribution is independent of θ, thus knowing θ does not yield us anything about the CDF we're interested in.