1 Basic concepts

1.1 Parametric vs Nonparametric models

A statistical model postulates a family of possible data generating mechanisms. Examples include:

(i) Let $X_1, \dots, X_n \sim^{\text{iid}} \Gamma(m, \theta)$ where m is known and $\theta \in (0, \infty) := \Theta$;

(ii) Let $Y_i = \alpha + \beta x_i + \varepsilon_i$ for $i \in [n] := \{1, \dots, n\}$, where $x_1, \dots, x_n$ and $\varepsilon_1, \dots, \varepsilon_n \sim^{\text{iid}} \mathcal{N}(0, \sigma^2)$. Here the unknown parameter is $\theta = (\alpha, \beta, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times (0, \infty) := \Theta$.

If the parameter space Θ is finite-dimensional, we speak of a *parametric model*. When the model is correctly specified, i.e there exists $\theta_0 \in \Theta$ for which the data were generated from the distribution with parameter θ_0, typically we can use the MLE $\hat{\theta}_n$ to estimate θ_0, and expect $n^{1/2}(\hat{\theta}_n - \theta_0)$ to converge to a non-degenerate limiting distribution. On the other hand, when the model is misspecified, inferences may be very misleading.

Examples of nonparametric models include:

(i) Let $X_1, \dots, X_n \sim^{\text{iid}} F$ for some unknown distribution function F;

(ii) Let $X_1, \dots, X_n \sim^{\text{iid}} f$ for some density f belonging to some unknown smoothness class;

(iii) Let $Y_i = m(x_i) + \varepsilon_i$ for $i \in [n]$, where $x_1, \dots, x_n$ are known, m belongs to some unknown smoothness class and $\varepsilon_1, \dots, \varepsilon_n$ are iid with $\mathbb{E}(\varepsilon_1) = 0$, $\mathrm{Var}(\varepsilon_1) = \sigma^2$.

Such infinite-dimensional models are much less vulnerable to model misspecification. Typically however, we will pay a price in terms of a slower rate of convergence.

1.2 Estimating an arbitrary distribution function

Let $\mathcal{F}$ denote the set of all distribution functions on $\mathbb{R}$. The *empirical distribution function* $\mathbb{F}_n$ of real-valued random variables $X_1, \dots, X_n$ is defined by

$$\mathbb{F}_n(x) = \mathbb{F}_n(x, X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \leq x\}.$$

Theorem (Glivenko-Cantelli Theorem). *Let $X_1, \dots, X_n \sim^{iid} F \in \mathcal{F}$ and let $\mathbb{F}_n$ denote the empirical distribution function of $X_1, \dots, X_n$. Then*

$$\sup_{x \in \mathbb{R}} |\mathbb{F}_n(x) - F(x)| \xrightarrow{a.s} 0 \text{ as } n \to \infty.$$

Proof. Let $\varepsilon > 0$ and $k := \lceil \frac{1}{\varepsilon} \rceil$. Let $x_0 = -\infty$, $x_i = \inf\{x \in \mathbb{R} : F(x) \geq i/k\}$ for $i \in [k-1]$ and $x_k = \infty$. Writing $F(x-)$ for $\lim_{y \uparrow x} F(y)$, note that for $i \in [k]$

$$F(x_i-) - F(x_{i-1}) \leq \frac{i}{k} - \frac{i-1}{k} = \frac{1}{k} \leq \varepsilon.$$

Now define the event

$$\Omega_{n,\varepsilon} = \left\{ \max_{i \in [k]} \sup_{m \geq n} |\mathbb{F}_m(x_i) - F(x_i)| \leq \varepsilon \right\} \cap \left\{ \max_{i \in [k]} \sup_{m \geq n} |\mathbb{F}_m(x_i-) - F(x_i-)| \leq \varepsilon \right\}$$

Noting that both $\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i \leq x\}$ and $\mathbb{F}_n(x-) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i < x\}$ are both sample averages of i.i.d random variables, we have by a union bound and the SLLN that

$$\begin{aligned}
&\mathbb{P}_F(\Omega_{n,\varepsilon}^c) \\
&\leq \sum_{i=1}^k \mathbb{P}_F\left(\sup_{m \geq n} |\mathbb{F}_m(x_i) - F(x_i)| > \varepsilon \right) + \sum_{i=1}^k \mathbb{P}_F\left(\sup_{m \geq n} |\mathbb{F}_m(x_i-) - F(x_i-)| > \varepsilon \right) \\
&\xrightarrow{a.s} 0.
\end{aligned}$$

Now let $x \in \mathbb{R}$ and find $i_* \in [k]$ such that $x \in [x_{i_*-1}, x_{i_*})$. Then for any $n_0 \in \mathbb{N}$ and $n \geq n_0$,

$$\begin{aligned}
\mathbb{F}_n(x) - F(x) &\leq \mathbb{F}_n(x_{i_*}-) - F(x_{i_*-1}) \\
&= \mathbb{F}_n(x_{i_*}-) - F(x_{i_*}-) + F(x_{i_*}-) - F(x_{i_*-1}) \\
&\leq \max_{i \in [k]} \sup_{m \geq n_0} |\mathbb{F}_m(x_i-) - F(x_i-)| + \varepsilon.
\end{aligned}$$

We also have

$$\begin{aligned}
F(x) - \mathbb{F}_n(x) &\leq F(x_{i_*}-) - \mathbb{F}_n(x_{i_*-1}) \\
&= F(x_{i_*}-) - F(x_{i_*-1}) + F(x_{i_*-1}) - \mathbb{F}_n(x_{i_*-1}) \\
&\leq \varepsilon + \max_{i \in [k]} \sup_{m \geq n_0} |\mathbb{F}_m(x_i) - F(x_i)|.
\end{aligned}$$

It follows that

$$\mathbb{P}_F\left(\sup_{n \geq n_0} \sup_{x \in \mathbb{R}} |\mathbb{F}_n(x) - F(x)| > 2\varepsilon \right) \leq \mathbb{P}_F(\Omega_{n_0,\varepsilon}^c) \to 0 \text{ as } n_0 \to \infty.$$

Since $\varepsilon > 0$ was arbitrary, we conclude that

$$\begin{aligned}
\mathbb{P}_F\left(\sup_{x \in \mathbb{R}} |\mathbb{F}_n(x) - F(x)| \to 0 \right) &= \mathbb{P}_F\left(\bigcap_{L=1}^\infty \bigcup_{n_0=1}^\infty \left\{ \sup_{n \geq n_0} \sup_{x \in \mathbb{R}} |\mathbb{F}_n(x) - F(x)| \leq \frac{1}{L} \right\} \right) \\
&= \lim_{L \to \infty} \lim_{n_0 \to \infty} \mathbb{P}_F\left(\sup_{n \geq n_0} \sup_{x \in \mathbb{R}} |\mathbb{F}_n(x) - F(x)| \leq \frac{1}{K} \right) \\
&= 1.
\end{aligned}$$

□

In fact, we can say much more:

Theorem (Dvoretzky-Kiefer-Wolfowitz Theorem)**.** *Under the conditions of the previous theorem, for every $\varepsilon > 0$,*

$$\mathbb{P}_F\left(\sup_{x\in\mathbb{R}} |\mathbb{F}_n(x) - F(x)| > \varepsilon\right) \leq 2e^{-2n\varepsilon^2}.$$

Proof. Not given. □

Corollary (Uniform Glivenko-Cantelli)**.** *Under the conditions of the Glivenko-Cantelli Theorem,*

$$\sup_{F\in\mathcal{F}} \mathbb{P}_F\left(\sup_{m\geq n}\sup_{x\in\mathbb{R}} |\mathbb{F}_m(x) - F(x)| > \varepsilon\right) \to 0 \ \text{as } n \to \infty.$$

Proof. By a union bound and the DKW inequality,

$$\begin{aligned}\sup_{F\in\mathcal{F}} \mathbb{P}_F\left(\sup_{m\geq n}\sup_{x\in\mathbb{R}} |\mathbb{F}_m(x) - F(x)| > \varepsilon\right) &\leq \sup_{F\in\mathcal{F}} \sum_{m\geq n} \mathbb{P}_F\left(\sup_{x\in\mathbb{R}} |\mathbb{F}_m(x) - F(x)| > \varepsilon\right) \\ &\leq 2\sum_{m\geq n} e^{-2m\varepsilon^2} \\ &\frac{2e^{-2n\varepsilon^2}}{1-e^{-2\varepsilon^2}} \to 0 \text{ as } n \to \infty.\end{aligned}$$

□

As another application of the DKW inequality, consider the problem of finding a confidence bound for F. Given $\alpha \in (0,1)$, set $\varepsilon_n = \varepsilon_n(\alpha) := \left\{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)\right\}^{1/2}$. Then, by the DKW Theorem,

$$\mathbb{P}_F\left(\max\{0, \mathbb{F}_n(x) - \varepsilon_n\} \leq F(x) \leq \min\{\mathbb{F}_n(x) + \varepsilon_n, 1\} \quad \forall x \in \mathbb{R}\right) \geq 1 - \alpha.$$

In fact, let $U_1, \dots, U_n \sim^{\text{iid}} \mathcal{U}[0,1]$ and let $\mathbb{G}_n$ denote their empirical distribution. Define the *quantile function* $F^{-1} : (0,1] \to (-\infty, \infty]$ by $F^{-1}(p) := \inf\{x \in \mathbb{R} : F(x) \geq p\}$ (i.e the *generalised inverse*). Since F is increasing and right-continuous, we have $\{x \in \mathbb{R} : F(x) \geq p\} = [F^{-1}(p), \infty)$, so $\{U_i \leq F(x)\} = \{F^{-1}(U_i) \leq x\}$. Hence

$$\begin{aligned}\mathbb{G}_n(F(x)) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{U_i \leq F(x)\} = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{F^{-1}(U_i) \leq x\} &=^d \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i \leq x\} \\ &= \mathbb{F}_n(x).\end{aligned}$$

So

$$\sup_{x\in\mathbb{R}} |\mathbb{F}_n(x) - F(x)| =^d \sup_{x\in\mathbb{R}} |\mathbb{G}_n(F(x)) - F(x)| \leq \sup_{t\in[0,1]} |\mathbb{G}_n(t) - t|$$

with equality if F is continuous. Thus if F is continuous, the distribution of $\sup_{x\in\mathbb{R}} |\mathbb{F}_n(x) - F(x)|$ does not depend on F!

Other generalisations of the Glivenko-Cantelli Theorem include Uniform Laws of Large Numbes (ULLN). Let $X_1, X_2, \ldots, X_n$ be iid, taking values in some measurable space $(\mathcal{X}, \mathcal{A})$, and let $\mathcal{G}$ denote a family of real-valued measurable functions on $\mathcal{X}$. We say $\mathcal{G}$ satisfies a *ULLN* if

$$\sup_{g\in\mathcal{G}} \left| \frac{1}{n}\sum_{i=1}^{n} g(X_i) - \mathbb{E}g(X) \right| \xrightarrow{a.s} 0.$$

Remark. $\mathcal{G}$ may be uncountable, so this random variable could be non measurable. But we can get around this by working with an outer probability $\mathbb{P}^*(A) = \inf\{\mathbb{P}(B) : B \in \mathcal{F},\ B \supseteq A\}$.

Thus, in the Glivenko-Cantelli Theorem we proved that $\{\mathbb{1}\{\cdot \leq x\} : x \in \mathbb{R}\}$ satisfies a ULLN. In general, proving a ULLN requires control of the 'size' of $\mathcal{G}$, which can be measured for instance through its metric entropy (see van de Geer, 2000).

1.3 Concentration Inequalities

Definition. A random variable X is *sub-Gaussian with variance parameter* σ^2 if

$$\mathbb{E}(e^{tX}) \leq e^{t^2\sigma^2/2} \quad \forall t \in \mathbb{R}.$$

Remark. It can be shown (see Example Sheet) that any such random variable must satisfy $\mathbb{E}(X) = 0$ and $\mathrm{Var}(X) \leq \sigma^2$. Note also that we have equality in the definition if $X \sim \mathcal{N}(0, \sigma^2)$.