Outline

1) Convergence in Probability and Distribution

2) Continuous Mapping, Slutsky's Theorem

3) Delta method

Asymptotics

[So far, everything has been finite-sample, often using special properties of model P (e.g. exp. fam.) to do exact calculations.]

[For "generic" models, exact calculations may be intractable or impossible. But we may be able to approximate our problem with a simpler problem in which calculations are easy]

[Typically approximate by Gaussian, by taking limit as # observations $\rightarrow \infty$. But this is only interesting if approx. is good for "reasonable" sample size.]

Convergence

Let $X_1, X_2, \ldots \in \mathbb{R}^d$ sequence of random vectors

We care about 2 kinds of convergence:

1) Cvg. in probability $(X_n \approx \text{constant})$
2) Cvg. in distribution $(X_n \approx N_d(0, I_d), \text{usually})$

We say the sequence __converges in probability__ to $c \in \mathbb{R}^d$ $(X_n \xrightarrow{P} c)$ if

$$\mathbb{P}(\|X_n - c\| > \varepsilon) \to 0, \quad \forall \varepsilon > 0$$

(could really be any distance on any $\mathcal{X}$)

[Can converge to a r.v. X too, but we don't need this]

We say the sequence __converges in distribution__ to random variable X $(X_n \Rightarrow X, X_n \xrightarrow{d} X)$ if

$$\mathbb{E} f(X_n) \to \mathbb{E} f(X) \quad \text{for all bdd, cts } f: \mathcal{X} \to \mathbb{R}$$

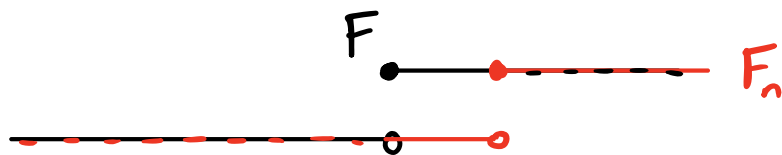__Thm__ $X_1, X_2, \ldots \in \mathbb{R}$, $F_n(x) = \mathbb{P}(X_n \leq x)$, $F(x) = \mathbb{P}(X \leq x)$

Then $X_n \Rightarrow X$ iff $F_n(x) \to F(x)$ $\forall x : F$ cts at x

Also known as __weak convergence__

Ex: If $X_n \sim \delta_{1/n}$, $\left(X_n \stackrel{a.s.}{=} \frac{1}{n} \right)$ $X \sim \delta_0$, then $X_n \Rightarrow X$

$$F_n(x) = 1\{\tfrac{1}{n} \le x\} \longrightarrow 1\{0 \le x\} \quad \underline{except} \quad x = 0$$

Prop $X_n \stackrel{p}{\longrightarrow} c$ iff $X_n \Rightarrow \delta_c$

Proof $(\Leftarrow)$ Let $f_\varepsilon(x) = \max\left(1, \frac{\|x-c\|}{\varepsilon}\right) \ge 1\{\|x-c\| > \varepsilon\}$

$$\mathbb{P}(\|X_n - c\| > \varepsilon) \le \mathbb{E}\, f_\varepsilon(X_n) \longrightarrow 0$$

$(\Rightarrow)$ f bdd, cts, note $\mathbb{E}\, f(X) = f(c)$

$\forall \varepsilon > 0$, $\exists d(\varepsilon) > 0$ s.t. $\|x - c\| \le d(\varepsilon) \Rightarrow |f(x) - f(c)| \le \varepsilon$

$$\mathbb{E}\, f(X_n) - f(c) \le \mathbb{E}\Big[\, |f(X_n) - f(c)| \cdot \big(1\{\|X_n - c\| \le d(\varepsilon)\} + 1\{\|\cdot\| > d\}\big)\Big]$$

$$\le \varepsilon + \mathbb{P}(\|X_n - c\| > d(\varepsilon)) \cdot \sup_x |f(x) - f(c)|$$

$$\le 2\varepsilon \cdot \sup |f| \quad \text{for suff. large } n \quad \boxtimes$$

In a sequence of statistical models $\mathcal{P}_n = \{P_{n,\theta} : \theta \in \Theta\}$

with $X_n \sim P_{n,\theta}$, we say $\delta_n(X_n)$ is $\underline{consistent}$

for $g(\theta)$ if $\delta_n(X_n) \stackrel{P_\theta}{\longrightarrow} g(\theta)$, meaning

$$\mathbb{P}_\theta\big(\|\delta_n(X_n) - g(\theta)\| > \varepsilon\big) \longrightarrow 0$$

Usually we omit the index n; sequence is implicit.

Limit Theorems

Let $X_1, X_2, \ldots$ iid random vectors

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Law of large numbers (LLN)

If $\mathbb{E}|X_i| < \infty$, $\mathbb{E}X_i = \mu$, then $\bar{X}_n \xrightarrow{P} \mu$ $\left(\bar{X}_n \xrightarrow{a.s.} \mu \right)$

Central limit theorem (CLT)

If $\mathbb{E}X = \mu \in \mathbb{R}^d$, $\text{Var}(X_n) = \Sigma$ (finite)

Then $\sqrt{n}(\bar{X}_n - \mu) \Rightarrow N(0, \Sigma)$

[There are stronger versions of both the LLN & CLT, but this will generally be enough for us]

Continuous Mapping

Theorem (Cts Mapping) g cts; $X_1, X_2, \ldots$ r.v.s

$$\text{If} \quad X_n \Rightarrow X \quad \text{then} \quad g(X_n) \Rightarrow g(X)$$

$$\text{If} \quad X_n \xrightarrow{P} c \quad \text{then} \quad g(X_n) \xrightarrow{P} g(c)$$

Proof f bdd, cts $\Rightarrow$ $f \circ g$ bdd, cts

$$\text{If} \quad X_n \Rightarrow X \quad \text{then} \quad \mathbb{E} f(g(X_n)) \longrightarrow \mathbb{E} f(g(X))$$

$X_n \xrightarrow{P} c$ special case with $X \sim \delta_c$ ⊠

Theorem (Slutsky) Assume $X_n \Rightarrow X$, $Y_n \xrightarrow{P} c$

Then: $\quad X_n + Y_n \Rightarrow X + c$

$$X_n \cdot Y_n \Rightarrow cX$$

$$X_n / Y_n \Rightarrow X/c \quad \text{if} \quad c \neq 0$$

Proof Show $(X_n, Y_n) \Rightarrow (X, c)$, apply cts mapping.

[Wouldn't normally be true that $X_n \Rightarrow X$, $Y_n \Rightarrow Y$
implies $(X_n, Y_n) \Rightarrow (X, Y)$ without specifying joint
dist.]

Theorem (Delta Method)

If
- $\sqrt{n}(X_n - \mu) \Rightarrow N(0, \sigma^2)$
- $f(x)$ differentiable at $x = \mu$

Then $\sqrt{n}\left(f(X_n) - f(\mu)\right) \Rightarrow N\left(0, \dot{f}(\mu)^2 \sigma^2\right)$

Informal statement:

$$X_n \approx N(\mu, \sigma^2/n) \implies f(X_n) \approx N\left(f(\mu), \dot{f}(\mu) \sigma^2/n\right)$$

Proof

$$f(X_n) = f(\mu) + \dot{f}(\mu)(X_n - \mu) + o(X_n - \mu)$$

$$\sqrt{n}\left(f(X_n) - f(\mu)\right) = \dot{f}(\mu) \cdot \sqrt{n}(X_n - \mu) + \sqrt{n} \cdot o(X_n - \mu)$$

$$= N\left(0, \dot{f}(\mu)^2 \sigma^2\right) \qquad \xrightarrow{P} 0$$

Multivariate:

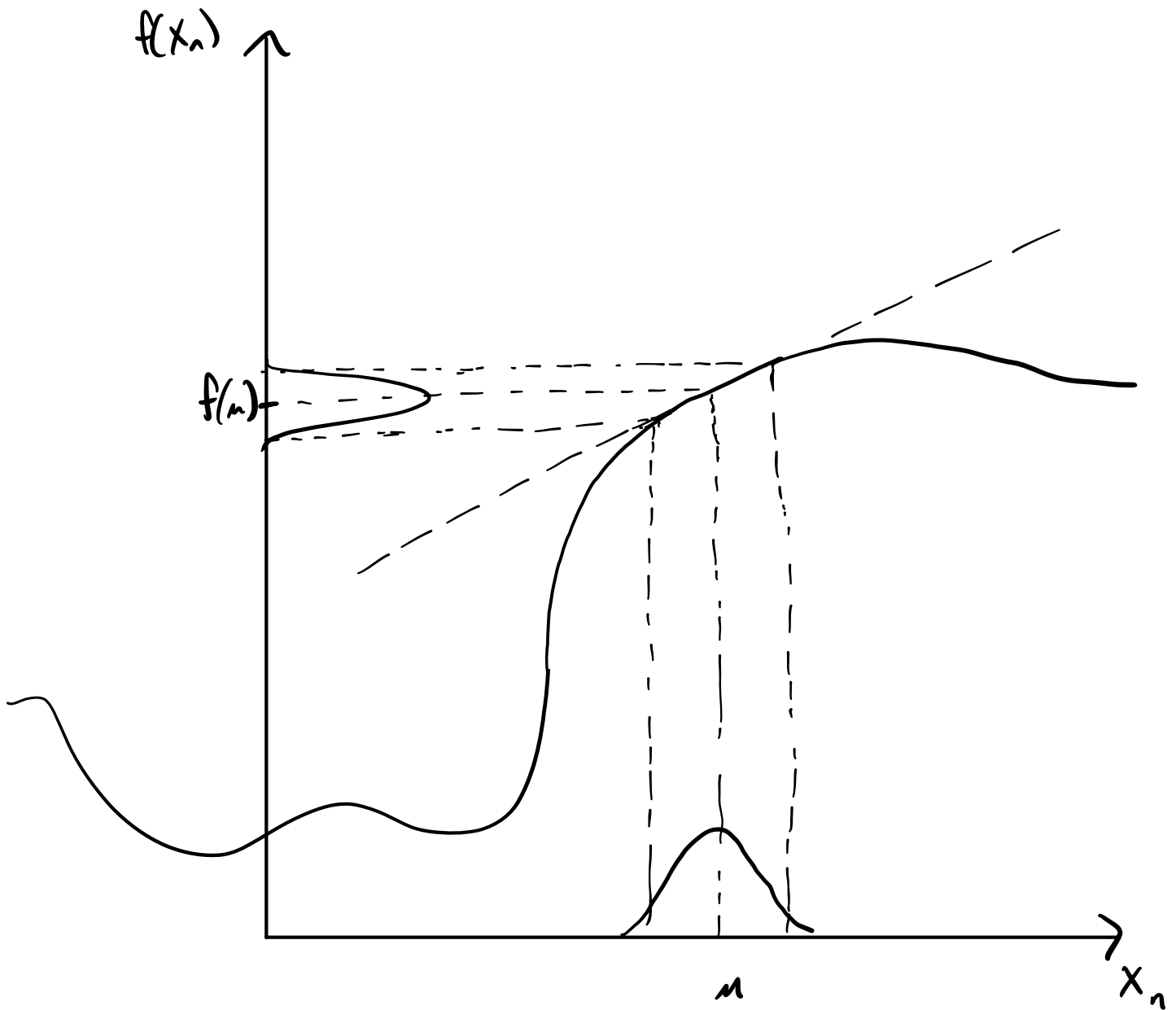$\sqrt{n}(X_n - \mu) \Rightarrow N_d(0, \Sigma)$, $f: \mathbb{R}^d \to \mathbb{R}^k$

Derivative $Df(x) = \begin{pmatrix} - \nabla f_1(x) - \\ \vdots \\ - \nabla f_k(x) - \end{pmatrix}$ exists at μ

Then $\sqrt{n}\left(f(X_n) - f(\mu)\right) \approx \sqrt{n} Df(\mu)(X_n - \mu)$

$$\approx N_k\left(0, Df(\mu) \Sigma Df(\mu)'\right)$$

$$= N\left(0, \nabla f(\mu)' \Sigma \nabla f(\mu)\right) \text{ if } k = 1$$

Delta Method

$$\left[\text{Scaling factor doesn't need to be } \sqrt{n}, \text{ but need } X_n - \mu \xrightarrow{P} 0 \right]$$

Ex $X_1, \ldots, X_n \overset{iid}{\sim} (\mu, \sigma^2)$

$Y_1, \ldots, Y_n \overset{iid}{\sim} (\upsilon, \tau^2)$ X, Y indep.

For large n, what is the distribution of $(\bar{X} + \bar{Y})^2$?

1) $\bar{X} \overset{P}{\to} \mu, \quad \bar{Y} \overset{P}{\to} \upsilon$ as $n \to \infty$

$\Rightarrow (\bar{X} + \bar{Y})^2 \overset{P}{\to} (\mu + \upsilon)^2$ ✓

2) $\sqrt{n}(\bar{X} - \mu) \Rightarrow N(0, \sigma^2)$ $\sqrt{n}(\bar{Y} - \upsilon) \Rightarrow N(0, \tau^2)$

Let $f(x, y) = (x + y)^2$

$\dfrac{\partial f}{\partial x}(x, y) = \dfrac{\partial f}{\partial y}(x, y) = 2(x + y)$

$f(\bar{x}, \bar{y}) \approx N\left(f(\mu, \upsilon), \; \nabla f' \begin{pmatrix} \sigma^2 & 0 \\ 0 & \tau^2 \end{pmatrix} \nabla f / n \right)$

$= N\left((\mu + \upsilon)^2, \; 4(\mu + \upsilon)^2(\sigma^2 + \tau^2)/n \right)$

More accurate:

$\sqrt{n}\left((\bar{x} + \bar{y})^2 - (\mu + \upsilon)^2 \right) \Rightarrow N\left(0, \; 4(\mu + \upsilon)^2(\sigma^2 + \tau^2) \right)$

3) What if $(\mu + \upsilon)^2 = 0$? Conclusion still holds:

$\sqrt{n}(\bar{X} + \bar{Y})^2 \overset{P}{\to} 0$

Note $\sqrt{n}\bar{X} + \sqrt{n}\bar{Y} \Rightarrow N(0, \sigma^2 + \tau^2)$ (cts mapping)

not **Slutsky !!**

So $n(\bar{X} + \bar{Y})^2 \Rightarrow (\sigma^2 + \tau^2)\chi_1^2$ (cts mapping)

why not delta method?

In general, can do higher-order Taylor expansions for delta method if derivatives $= 0$:

$$f(X_n) \approx \underbrace{f(m)}_{O(1)} + \underbrace{\dot{f}(m)(X_n - m)}_{O_p(n^{-1/2})} + \underbrace{\frac{\ddot{f}(m)}{2}(X_n - m)^2}_{O_p(n^{-1})} + \cdots$$

If $\dot{f}(m) = 0$, use second-order term:

$$n\left(f(X_n) - f(m)\right) \approx \frac{\ddot{f}(m)}{2}\left(\sqrt{n}(X_n - m)\right)^2$$

$$\approx \frac{\ddot{f}(m)\sigma^2}{2}\chi_1^2$$