

Outline

10/28/2021

- 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 \mathcal{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 $X_1, X_2, \dots \in \mathbb{R}^d$ sequence of random vectors

We care about 2 kinds of convergence:

1) cvg. in probability ($X_n \approx \text{constant}$)

2) cvg. in distribution ($X_n \approx N_d(0, I_d)$, usually)

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

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

(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

$$\mathbb{E} f(X_n) \rightarrow \mathbb{E} f(X) \quad \text{for all bdd, cts } f$$

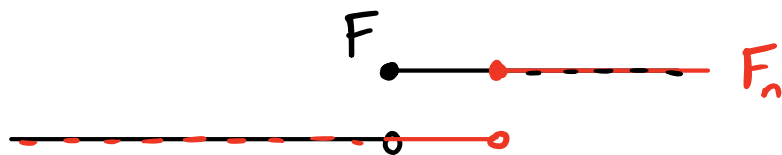
Thm $X_1, X_2, \dots \in \mathbb{R}, F_n(x) = \mathbb{P}(X_n \leq x), F(x) = \mathbb{P}(X \leq x)$

Then $X_n \Rightarrow X$ iff $F_n(x) \rightarrow F(x) \quad \forall x: F \text{ cts at } x$

Also known as weak convergence

Ex: If $X_n \sim \delta_n^{(x_n = \frac{1}{n})}$, $X \sim \delta_0$, then $X_n \Rightarrow X$

$$F_n(x) = 1\{\frac{1}{n} \leq x\} \rightarrow 1\{0 \leq x\} \quad \text{except } x=0$$



Prop $X_n \xrightarrow{P} c$ iff $X_n \Rightarrow \delta_c$

Proof (\Leftarrow) Let $f_\varepsilon(x) = \max(1, \|x-c\|/\varepsilon) \geq 1\{\|x-c\| > \varepsilon\}$

$$P(\|X_n - c\| > \varepsilon) \leq E f_\varepsilon(X_n) \rightarrow 0$$

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

$$\forall \varepsilon > 0, \exists \delta_\varepsilon > 0 \text{ s.t. } \|x - c\| \leq \delta_\varepsilon \Rightarrow |f(x) - f(c)| \leq \varepsilon$$

$$\begin{aligned} E f(X_n) - f(c) &\leq E |f(X_n) - f(c)| \\ &\leq \varepsilon + P(\|X_n - c\| > \delta_\varepsilon) \cdot \sup_x |f(x) - f(c)| \\ &\leq 2\varepsilon \cdot \sup |f| \quad \text{for suff. large } n \quad \square \end{aligned}$$

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 consistent for $g(\theta)$ if $\delta_n(X_n) \xrightarrow{P_\theta} g(\theta)$, meaning

$$P_\theta(\|\delta_n(X_n) - g(\theta)\| > \varepsilon) \rightarrow 0$$

Usually we omit the index n ; sequence is implicit.

Limit Theorems

Let X_1, X_2, \dots 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 $\bar{X}_n \xrightarrow{P} \mu$ ($\bar{X}_n \xrightarrow{a.s.} \mu$)

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, \dots r.v.s

If $X_n \Rightarrow X$ then $g(X_n) \Rightarrow g(X)$

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

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

If $X_n \Rightarrow X$ then $\mathbb{E} f(g(X_n)) \rightarrow \mathbb{E} f(g(X))$

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

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

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

$X_n \cdot Y_n \Rightarrow cX$

$X_n / Y_n \Rightarrow X/c$ if $c \neq 0$

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)

$$\text{If } \sqrt{n}(X_n - \mu) \Rightarrow N(0, \sigma^2)$$

• $f(x)$ differentiable at $x = \mu$

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

Informal statement:

$$X_n \approx N(\mu, \sigma^2/n) \Rightarrow f(X_n) \approx N(f(\mu), f'(\mu)^2 \sigma^2/n)$$

Proof $f(X_n) = f(\mu) + f'(\mu)(X_n - \mu) + o(X_n - \mu)$

$$\begin{aligned} \sqrt{n}(f(X_n) - f(\mu)) &= f'(\mu) \cdot \sqrt{n}(X_n - \mu) + \underbrace{\sqrt{n} \cdot o(X_n - \mu)}_{\xrightarrow{P} 0} \\ &= N(0, f'(\mu)^2 \sigma^2) \end{aligned}$$

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

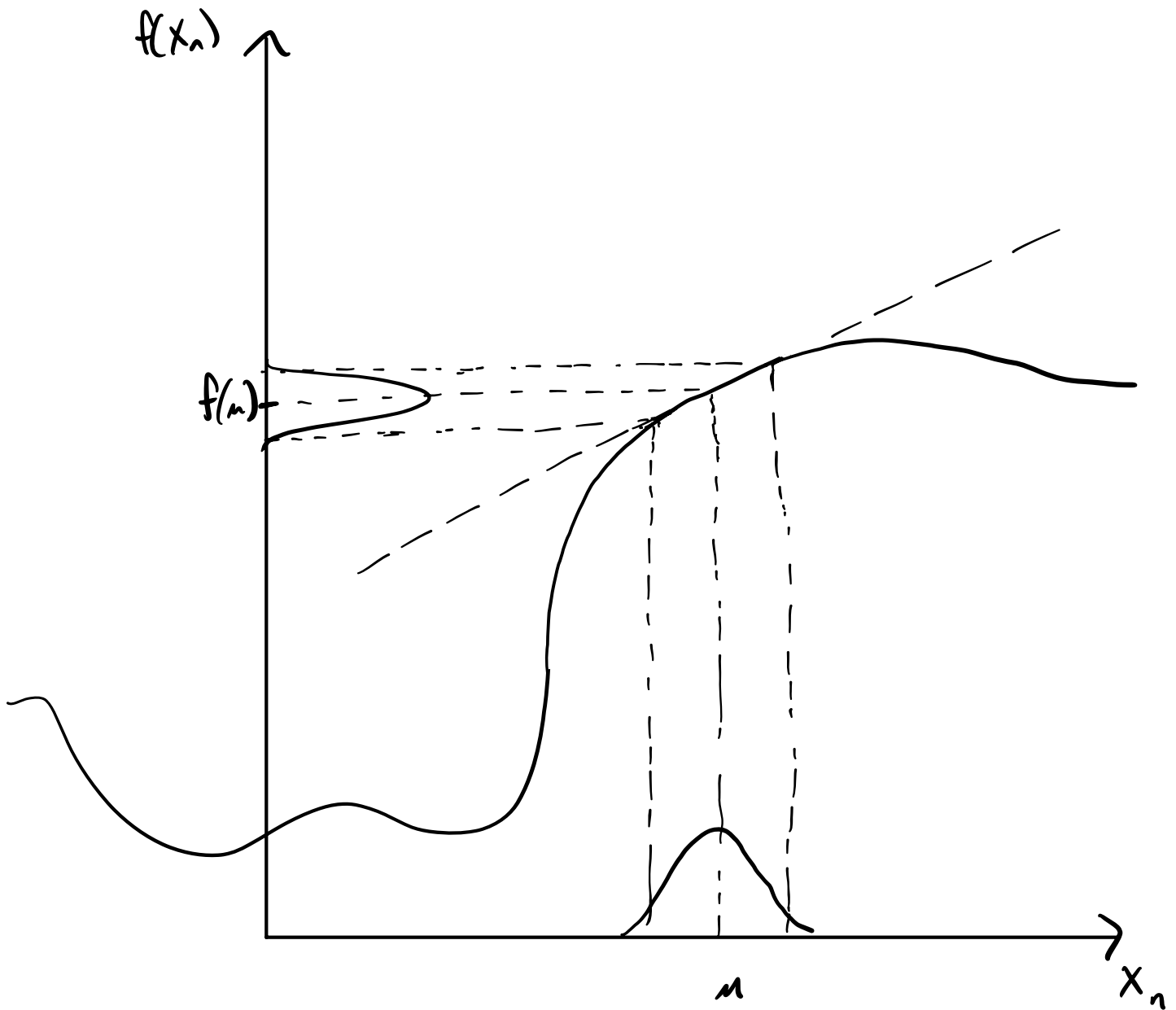
Derivative $Df(x) = \begin{pmatrix} -\nabla f_1(x) - \\ \vdots \\ -\nabla f_k(x) - \end{pmatrix}$ exists at μ

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

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

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

Delta Method



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

Ex $X_1, \dots, X_n \stackrel{iid}{\sim} (\mu, \sigma^2)$
 $Y_1, \dots, Y_n \stackrel{iid}{\sim} (\nu, \tau^2)$ X, Y indep.

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

1) $\bar{X} \xrightarrow{P} \mu, \quad \bar{Y} \xrightarrow{P} \nu \quad \text{as } n \rightarrow \infty$

$\Rightarrow (\bar{X} + \bar{Y})^2 \xrightarrow{P} (\mu + \nu)^2 \quad \checkmark$

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

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

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

$f(\bar{X}, \bar{Y}) \approx N(f(\mu, \nu), \nabla f' \begin{pmatrix} \sigma^2 & 0 \\ 0 & \tau^2 \end{pmatrix} \nabla f / n)$

$= N((\mu + \nu)^2, 4(\mu + \nu)^2(\sigma^2 + \tau^2) / n)$

More accurate:

$\sqrt{n}((\bar{X} + \bar{Y})^2 - (\mu + \nu)^2) \Rightarrow N(0, 4(\mu + \nu)^2(\sigma^2 + \tau^2))$

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

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

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 $\neq 0$:

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

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

$$\begin{aligned} n(f(X_n) - f(\mu)) &\approx \frac{\ddot{f}(\mu)}{2} (\sqrt{n}(X_n - \mu))^2 \\ &\approx \frac{\ddot{f}(\mu)\sigma^2}{2} \chi_1^2 \end{aligned}$$