Week 7

* the moments **do not** capture all info abt the dist. in the unbounded support scenario

Q: why do we care?

A: lies in need to approximate RV's

> **Thm (convergence of MGF's leads to convergence of CDF).**
> Suppose $X_1, X_2, \dots$ is a sequence of random variables, each w/ mgf $M_{X_i}(t)$.
> Suppose
> $$\lim_{i \to \infty} M_{X_i}(t) \to M_X(t)$$
> for all t in a neighborhood of 0, and $M_X(t)$ a mgf. (*)
> Then $X_i \to X$ in distribution
> where X is a RV w/ mgf $M_X(t)$
> i.e. $F_{X_i}(x) \to F_X(x)$ at all points x where cdf F_X is continuous

(*) $\exists\, h>0$ s.t. $\forall\, t \in (-h,h)$, $M_{X_i}(t) = M_X(t)$

Proof (idea). Beyond scope of class

> **Def**: $X_i \to X$ in distribution ($X_i \xrightarrow{d} X$)
> if $F_{X_i}(x) \to F_X(x)$ at all continuity points of F_X

- implicit in above thm.

Ex (Poisson Approx.): Let $X \sim \text{Binomial}(n,p)$. We know $\begin{cases} E(X) = np \\ V(X) = np(1-p) \end{cases}$

As n gets large, X "behaves" like a Poisson RV.

$Y \sim \text{Poisson}(\lambda)$ if
$$f_Y(y) = e^{-\lambda} \frac{\lambda^y}{y!}, \quad y = 0,1,\dots$$
$$E(Y) = V(Y) = \lambda$$
$$M_Y(t) = e^{\lambda(e^t - 1)}.$$

We will establish that
$$P(X \le x) \simeq P(Y \le x) \text{ as } n \uparrow \infty \ \forall\, x \in \mathbb{N}.$$

More precisely, we also need $p \downarrow 0$
s.t. $np = \text{constant} \equiv \lambda$.

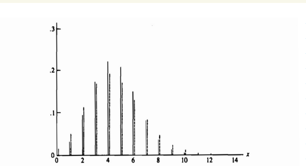


Figure 2.3.3. *Poisson (dotted line) approximation to the binomial (solid line), $n = 15$, $p = .3$*

Suppose $\begin{cases} np = \lambda = \text{const} \\ n \uparrow \infty,\ p \downarrow 0. \end{cases}$

$M_X(t) = (pe^t + 1 - p)^n$

since $X_n \overset{d}{=} Y_1 + \dots + Y_n$ where $Y_i \overset{iid}{\sim} \text{Bernoulli}(p)$ s.t. $\begin{cases} P(Y_i = 1) = p \\ P(Y_i = 0) = 1-p \end{cases}$

$\Rightarrow M_{X_n}(t) = E(e^{tX_n}) = E(e^{t(Y_1 + \dots + Y_n)}) = E(e^{tY_1}) \cdots E(e^{tY_n})$

and $Y_1, \dots, Y_n$ are independent
$$= \prod_{i=1}^{n} E(e^{tY_i}) = (pe^t + 1 - p)$$

so, $M_{X_n}(t) = (pe^t + (1-p))^n$
$$= (1 + p(e^t - 1))^n$$
$$= \left(1 + \tfrac{1}{n}(e^t - 1)np\right)^n, \quad p = \tfrac{1}{n}np$$
$$= \left(1 + \tfrac{1}{n}(e^t - 1)\lambda\right)^n$$
$$= \left(1 + \tfrac{1}{n}(e^t - 1)\lambda\right)^n \quad \leftarrow \text{what?}$$
$$= \left(1 + \tfrac{1}{n}(e^t - 1)\lambda\right)^{\frac{n}{(e^t-1)\lambda}(e^t - 1)\lambda}, \quad e = \lim_{x \to \infty}\left(1 + \tfrac{1}{x}\right)^x$$
$$\to e^{(e^t - 1)\lambda} \text{ as } n \to \infty$$

Hence, $M_X(t) \to M_Y(t)\ \forall\, t$

By the convergence thm. of mgf,
if $X_n \sim \text{Binomial}(n,p)$, $\lambda = np$. then
$$F_{X_n}(x) \to F_Y(x) \text{ as } n \to \infty \ \forall\, x \text{ where } F_Y \text{ is continuous.}$$

For Poisson, F_Y is a step function w/ continuity at $y \notin \mathbb{N}$.

For $y \in \mathbb{N}$, $F_Y(y) = F_Y(y + \frac{1}{2})$ continuous at $y + \frac{1}{2}$.

so,
$P(X_n \le x) \to P(Y \le x)$ as $n \to \infty\ \forall\, x \in \mathbb{N}$ (in fact, true $\forall\, x \in \mathbb{R}$).

Remark: still holds if $X_n \sim \text{Binomial}(n, p_n)$ and $n \to \infty$, $p_n \to 0$ s.t. $np_n \to \lambda$.

Thm. $\forall a, b \in \mathbb{R}$,

$$\boxed{M_{aX+b}(t) = e^{bt} M_X(at)}$$

Proof (Sketch).

$$M_{aX+b} = E[e^{t(aX+b)}] = E[e^{taX} e^{bt}] = e^{bt} E[e^{taX}] = e^{bt} M_X(at). \square$$

2.4: Tools

Interchange Integral and Differential

Thm (Leibniz's Rule). If

- $f(x, \theta), a(\theta), b(\theta)$ are differentiable wrt θ
- $\frac{\partial f}{\partial \theta}(x, \theta)$ is continuous on $\underbrace{x \in (x_0, x_1)}_{\text{can be weakened}}$

Then for $a(\theta), b(\theta) \in (x_0, x_1)$:

$$\boxed{\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} f(x, \theta)\, dx = f(b(\theta), \theta) \frac{d}{d\theta} b(\theta) - f(a(\theta), \theta) \frac{d}{d\theta} a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial \theta} f(x, \theta)\, dx}$$

Corollary: If $a(\theta) = a$, $b(\theta) = b$, then

$$\frac{d}{d\theta} \int_a^b f(x, \theta)\, dx = \int_a^b \frac{\partial f}{\partial \theta}(x, \theta)\, dx$$

Proof (Leibniz's Rule).

$$\text{LHS} = \lim_{\delta \to 0} \frac{1}{\delta} \left(\int_{a(\theta+\delta)}^{b(\theta+\delta)} f(x, \theta+\delta)\, dx - \int_{a(\theta)}^{b(\theta)} f(x, \theta)\, dx \right)$$

$$= \lim_{\delta \to 0} \frac{1}{\delta} \Big[\underbrace{\int_{a(\theta)}^{b(\theta)} [f(x, \theta+\delta) - f(x, \theta)]\, dx}_{:= A} + \underbrace{\int_{a(\theta+\delta)}^{a(\theta)} f(x, \theta+\delta)\, dx}_{:= B} + \underbrace{\int_{b(\theta)}^{b(\theta+\delta)} f(x, \theta+\delta)\, dx}_{:= C} \Big]$$

where

$$A = \lim_{\delta \to 0} \frac{1}{\delta} \int_{a(\theta)}^{b(\theta)} f(x, \theta+\delta) - f(x, \theta)\, dx$$

$$= \int_{a(\theta)}^{b(\theta)} \lim_{\delta \to 0} \frac{1}{\delta} [f(x, \theta+\delta) - f(x, \theta)]\, dx$$

$$= \int_{a(\theta)}^{b(\theta)} \frac{\partial f}{\partial \theta}(x, \theta)\, d\theta \quad \text{since} \begin{cases} \frac{\partial f}{\partial \theta} \text{ is cont. in } x \\ \text{next theorem} \end{cases}$$

$$B = \lim_{\delta \to 0} -\frac{1}{\delta} \int_{a(\theta)}^{a(\theta+\delta)} f(x, \theta+\delta)\, dx$$

$$= \lim_{\delta \to 0} -\frac{a(\theta+\delta) - a(\theta)}{\delta} \times \frac{1}{a(\theta+\delta) - a(\theta)} \int_{a(\theta)}^{a(\theta+\delta)} f(x, \theta+\delta)\, dx$$

$$= -\frac{d}{d\theta} a(\theta) \times f(a(\theta), \theta)$$

Likewise,

$$C = \frac{d}{d\theta} b(\theta) f(b(\theta), \theta). \square$$

To strengthen the above result for infinite domain of the integral, we need another tool:

Interchanging Limit and Integral

Thm. Suppose

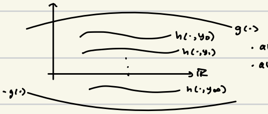
- $h(x, y)$ is continuous at $y = y_0$ for each fixed x
- there is a function $g(x)$ s.t.

$$\text{envelope condition} \begin{cases} |h(x, y)| \le g(x) \quad \forall x, y \\ \int_{-\infty}^{\infty} g(x)\, dx < \infty \end{cases}$$

Then,

$$\boxed{\lim_{y \to y_0} \int_{-\infty}^{\infty} h(x, y)\, dx = \int_{-\infty}^{\infty} \lim_{y \to y_0} h(x, y)\, dx}$$

Remarks:
- g is called the dominating or envelope function of f
- known as (a version/consequence of) Lebesgue's Dominated Convergence Thm (DCT)
- think of h as a collection of functions in x $\{h(x, y)\}_{y \in \mathbb{R}}$



- all $h(x, y)$ must be contained in $(-g, g)$
- all g must be integrable

Interchange Differential and Integral

Thm. Suppose

- For each x, $f(x, \theta)$ is differentiable wrt θ, at $\theta \in \Theta$
- For each $\theta \in \Theta$, there is a function $g(\cdot, \theta)$ and $\delta_0 > 0$ s.t.

$$\begin{cases} \left| \frac{\partial}{\partial \theta} f(x, \theta) \big|_{\theta = \theta'} \right| \le g(x, \theta') \quad \forall \theta' \in (\theta - \delta_0, \theta + \delta_0) \\ \int_{-\infty}^{\infty} g(x, \theta)\, dx < \infty \quad \text{for each } \theta \in \Theta \end{cases}$$

Then

$$\boxed{\frac{d}{d\theta} \int_{-\infty}^{\infty} f(x, \theta)\, dx = \int_{-\infty}^{\infty} \frac{d}{d\theta} f(x, \theta)\, dx}$$

holds for each $\theta \in \Theta$

Remarks:
- Leibniz's rule holds under this weaker envelope condition for $f(x, \theta)$.
- If we need only to differentiate at $\theta = \theta_0$, then it is sufficient that the envelope condition be satisfied for a neighborhood of θ_0 i.e. $\forall \theta \in (\theta_0 - \delta_0, \theta + \delta_0)$ for some $\delta_0 > 0$, and only a dominating function $g(x, \theta_0)$ (for fixed $\theta = \theta_0$) is required
- This Thm is a direct consequence of Lebesgue's Dominated Convergence Thm (DCT)

<u>Ex(Exponential)</u>: Let $X \sim EXP(\lambda)$; $f(x) = \frac{1}{\lambda} e^{-x/\lambda}$, $\lambda \in [0, \infty)$.

Moments: $E(X^n) = \int_0^\infty \frac{1}{\lambda} x^n e^{-x/\lambda} dx$, $n = 1, 2, \ldots$

wish to differentiate

$$\frac{d}{dx} E(X^n) = \frac{d}{dx} \int_0^\infty \cdot = \int_0^\infty \frac{d}{dx} \cdot \quad \text{, need to verify } (*)$$

$$= \int_0^\infty x^n \left(\frac{-1}{\lambda^2} e^{-x/\lambda} + \frac{1}{\lambda} e^{-x/\lambda} \frac{x}{\lambda^2} \right) dx$$

$$= \int_0^\infty x^n e^{-x/\lambda} \frac{1}{\lambda^2} \left(\frac{x}{\lambda} - 1 \right) dx$$

$$= \int_0^\infty x^{n+1} e^{-x/\lambda} \frac{1}{\lambda^3} - x^n e^{-x/\lambda} \frac{1}{\lambda^2} dx$$

$$= \frac{1}{\lambda^2} E(X^{n+1}) - \frac{1}{\lambda} E(X^n)$$

This gives a recursive relation, i.e.

$$\underline{E(X^{n+1}) = \lambda E(X^n) + \lambda^2 \frac{d}{dx} E(X^n)}$$

(Similar identity holds for broad family of dist.'s)

Now, we justify the interchange of $\int$ and d/dx, i.e. $(*)$

where $\frac{d}{dx} f(x, \lambda) = x^n e^{-x/\lambda} \frac{1}{\lambda^2} \left(\frac{x}{\lambda} - 1 \right)$.

Need to find a dominating function $g(\cdot, \lambda)$ s.t.

$$\begin{cases} 1) & \left| \frac{d}{d\theta} f(x, \theta) \big|_{\theta = \theta'} \right| < g(x, \lambda) \quad \forall \theta' \in (\lambda - \delta_0, \lambda + \delta_0) \\ 2) & g \text{ integrable wrt } x \end{cases}$$

$$\left| \frac{d}{d\theta} f(x, \theta) \right| = x^n e^{-x/\theta} \frac{1}{\theta^2} \left| \frac{x}{\theta} - 1 \right|$$

$$\leq x^n e^{-x/\theta} \frac{1}{\theta^2} \left(\frac{x}{\theta} + 1 \right) \quad \text{since } x > 0$$

$$\leq x^n e^{-x/(\theta + \delta_0)} \frac{1}{(\theta - \delta_0)^2} \left(\frac{x}{\theta - \delta_0} + 1 \right) =: g(x, \theta)$$

where the last inequality holds $\forall x > 0$, $\forall \theta > \delta_0$.

Thus, ① holds for the chosen $g(x, \lambda)$.

To verify ②,

$$\int g(x, \lambda) dx = \int x^{n+1} e^{-x/\lambda + \delta_0} \frac{1}{(\lambda - \delta_0)^3} + \int x^n e^{-x/\lambda + \delta_0} \frac{1}{(\lambda - \delta_0)^2} dx$$

= Multiple of $(n+1)$-th moment of an exponential RV

+ multiple of n-th moment of another exp. RV

+ ... + " " " " $< +\infty$. □

<u>Ex(Gaussian)</u>: Let $X \sim N(0,1)$; $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $x \in \mathbb{R}$.

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^\infty e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

where $\frac{d}{dt} M_X(t) = \frac{d}{dt} E(e^{tX}) \overset{(*)}{=} E\left[\frac{d}{dt} (e^{tX}) \right]$

Justifying $(*)$ by finding dominating function for

$$\left| \frac{d}{dt} e^{tx} e^{-x^2/2} \right| = \left| x e^{tx} e^{-x^2/2} \right| = |x| e^{tx} e^{-x^2/2}$$

Let $0 < \delta_0 < t$:

$$\left.\begin{array}{l} \text{if } x \geq 0: |x| e^{tx} e^{-x^2/2} \leq x e^{(t + \delta_0)x} e^{-x^2/2} \\ \text{if } x < 0: |x| e^{tx} e^{-x^2/2} \leq (-x) e^{(t - \delta_0)x} e^{-x^2/2} \end{array}\right\} g(x, t)$$

Take the RHS as $g(x, t)$ for $x \geq 0$ and $x < 0$.

Then, $\left| \frac{d}{dt} e^{tx} e^{-x^2/2} \right|_{t = t'} \leq g(x, t) \quad \forall t' \in (t - \delta_0, t + \delta_0)$, i.e.

the domination holds.

Then, can check that $g(\cdot, t)$ is integrable:

$$\int_0 x e^{(t + \delta_0)x} e^{-x^2/2} + \int_{-\infty} (-x) e^{(t - \delta_0)x} e^{-x^2/2}$$

$< \infty$. □

<u>Interchanging Sum and Differential</u>

> Suppose $\sum_{x=0}^\infty h(x, \theta)$ exists (i.e. converges pointwise) $\forall \theta \in (a, b)$
>
> Moreover, assume
>
> 1) $\frac{\partial}{\partial \theta} h(x, \theta)$ is continuous in θ for each x
> 2) $\sum_{x=0}^\infty \frac{\partial}{\partial \theta} h(x, \theta)$ converges uniformly for all θ in a closed subinterval of (a, b)
>
> Then $\boxed{\frac{d}{d\theta} \sum_{x=0}^\infty h(x, \theta) = \sum_{x=0}^\infty \frac{\partial}{\partial \theta} h(x, \theta)}$

<u>Remark</u>: This is a consequence of Lebesgue's D.C.T

<u>Ex(Geometric)</u>: Let $X \sim \text{Geometric}(\theta)$, $0 < \theta < 1$

$P(X = x) = \theta (1 - \theta)^x$, $x = 0, 1, 2, \ldots$

This is a valid pmf since

$$\sum_{x=0}^\infty \theta (1 - \theta)^x = 1$$

Differentiating both sides wrt θ, assuming we can interchange Σ and $\frac{d}{d\theta}$:

$$\sum_{x=0}^\infty (1 - \theta)^x - \theta x (1 - \theta)^{x-1} = 0$$