Using the Berry-Esseen bound for sums of independent random variables (cf., e.g., Shorack (2000), page 259), we find, since ε_K are iid and $E|\frac{\varepsilon_K}{\sigma}|^3 = E|\frac{\varepsilon_1}{\sigma}|^3 = \tau_3$, for all k

$$
\begin{aligned}
\|G_{ni} - \Phi\|_\infty &\le 9\frac{\sum_{k=1}^{n} \mathbb{E}|a_k \varepsilon_k/\sigma|^3}{\left(\sum_{k=1}^{n} a_k^2\right)^{3/2}} = \frac{9\tau_3 \sum_{k=1}^{n} |a_k| \cdot |a_k|^2}{1} \\
&\le 9\tau_3 \max_{1\le k\le n} |a_k| \\
&\le 9\tau_3 \max_{1\le k\le n} \left(h_{kk}\right)^{1/2}.
\end{aligned}
$$

Lectures 28 - 29

Characteristic Functions

(Uniqueness and Inversion Theorems).

Definition

Let X be a r.v. with df F.

The characteristic function (ch) is:

$$\Phi(t) \equiv \Phi_X(t) \equiv E\, e^{itX}$$

$$= \int_{-\infty}^{\infty} e^{itx}\, dF(x)$$

$$= \int_{-\infty}^{\infty} \cos(tx)\, dF(x) + i \int_{-\infty}^{\infty} \sin(tx)\, dF(x)$$

$|z| = |a + ib| = \sqrt{a^2 + b^2}$

$e^{ia} = \cos(a) + i\sin(a)$

Remarks:

1) $\Phi(t)$ is, in general, a complex number
2) $\Phi(t)$ is defined for all $t \in \mathbb{R}$.

Properties of the ch Φ

1) $\Phi(0) = 1$
2) $|\Phi(t)| \leq 1$, for all $t \in \mathbb{R}$
(Recall that $|\cdot|$ here is the modulus of a complex number. If $z = a + ib$, then $|z| = \sqrt{a^2+b^2}$).
3) $\Phi_{aX+b}(t) = e^{itb}\,\Phi_X(at)$, for all $t \in \mathbb{R}$.
4) $\Phi(-t) = \bar{\Phi}(t)$, for all $t \in \mathbb{R}$
(If $z \in \mathbb{C}$, $z = a + ib$, we denote by $\bar{z} = a - ib$

5) If $X_1, \dots, X_n$ are independent

$$\Phi_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n \Phi_{X_i}(t)$$

6) Φ is real valued iff $X \cong -X$.

7) If X and X' are iid with chf Φ, then $X^s \equiv X - X'$ has chf $|\Phi|^2$.

8) $\Phi(\cdot)$ is uniformly continuous on $\mathbb{R}$.
(in tbk)

Proof:

7) $$\Phi_{X-X'} \underset{(5)}{=} \Phi_X \cdot \Phi_{-X} =$$

$$= \Phi_X(t) \cdot \Phi_X(-t) = \Phi_X \cdot \overline{\Phi}_X = |\Phi|^2.$$

8) „$\Leftarrow$" If $X \cong -X \Rightarrow \Phi_X = \Phi_{-X}$

But $\Phi_{-X} = \overline{\Phi}_X$ $\Bigg\} \Rightarrow$

$\Phi_X = \overline{\Phi}_X \Rightarrow \Phi_X$ is real.

„$\Rightarrow$" If Φ is real, then $\Phi_X = \overline{\Phi}_X = \Phi_{-X} \Rightarrow$

(By the uniqueness Theorem 2.1 page 346) $X \cong -X$.

Examples

1) If $X \sim f(x)$; $f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$ (Cauchy).

$$\Phi_X(t) = \int_{-\infty}^{\infty} e^{itx} \cdot \frac{1}{\pi} \cdot \frac{1}{1+x^2}\, dx$$

$$= e^{-t}.$$

2) If $X \sim N(0,1)$; then

$$\Phi_X(t) = e^{-\frac{t^2}{2}}.$$

[Computing the above integrals requires standard complex analysis tools such as the Residue Theorem and the Uniqueness Theorem of Analytic Functions. We do not present such details here.].

- In what follows we show that there exists a one-to-one correspondence between **ANY** F and Φ.
- The important thing to remember is that F does **NOT** have to have a density wrt the Lebesgue measure. Therefore the results presented below are general.
  See pages 5'-7' for more explanations.

• We first introduce the notion of a $\boxed{\text{complementary pair}}$ of r.vs.

Let $U \sim f_U(\cdot)$, continuous.

Let W denote a r.v. with bounded and continuous density $f_W(\cdot)$ and with chf $\Phi_W(\cdot)$ <u>such that</u>:

(2) $f_U(t) = c\,\Phi_W(-t)$, for all t and some constant c.

• Such pairs exist. One example is

$U \sim N(0,1)$
$W \sim N(0,1)$ and $c = \frac{1}{\sqrt{2\pi}}$.

(For more examples, see page 346).

Theorem 1 (Uniqueness)

Every df on the line (F) has a unique chf (Φ).

Theorem 2 (Inversion ← mostly qualitative)

Let X be a r.v. with df $F_X(\cdot)$ and chf $\Phi_X(\cdot)$, we always have:

$$F_X(r_2) - F_X(r_1) = \lim_{a \to 0} \int_{r_1}^{r_2} f_a(t)\,dt, \quad \text{for all } r_1 < r_2 \text{ cont. points of } F. \quad (3)$$

Here $f_a(\cdot)$ is the density of the r.v.

$X_a \equiv X + aU$, with U given as in (2) above.

and (4) $$f_a(t) = \int_{-\infty}^{\infty} e^{-itv}\,\Phi_X(v) \cdot c\, f_W(av)\,dv,$$

for all $t \in \mathbb{R}$.

Remark Note that ③ and ④ together establish that there is a unique correspondence between F_X and Φ_X.

(Not required for exams)

Proof: By the convolution formula (*) page 3 and a change of variables we have:

$$f_a(t) = \int_{-\infty}^{\infty} \frac{1}{a} f_V\left(\frac{t-x}{a}\right) dF_X(x)$$

$$\underset{\substack{\uparrow \\ (2)}}{=} \frac{c}{a} \int_{-\infty}^{\infty} \Phi_W\left(\frac{x-t}{a}\right) dF_X(x)$$

$$\underset{\substack{\uparrow \\ \text{Definition of } \Phi_W}}{=} \frac{c}{a} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{i(x-t)w/a} f_W(w)\, dw \right) dF_X(x)$$

$$\underset{\substack{\uparrow \\ \text{Fubini}}}{=} \frac{c}{a} \int_{-\infty}^{\infty} e^{-itw/a} f_W(w) \left(\int_{-\infty}^{\infty} e^{ixw/a}\, dF_X(x) \right) dw$$

$$\underset{\substack{\uparrow \\ \text{Definition of} \\ \Phi_X}}{=} \frac{c}{a} \int_{-\infty}^{\infty} e^{-itw/a} \, \Phi_X(w/a) \, f_W(w) \, dv$$

$$\underset{\substack{\uparrow \\ \text{change} \\ \text{of variables}}}{=} c \int_{-\infty}^{\infty} e^{-itv} \, \Phi_X(v) \, f_W(av) \, dv$$

Thus:

$$f_a(t) = c \int_{-\infty}^{\infty} e^{-itv} \, \Phi_X(v) \, f_W(av) \, dv,$$ as claimed.

Thus, since $X_a \xrightarrow{d} X$:

$$F_X(r_2) - F_X(r_1) = \lim_{a \to \infty} \left(F_a(r_2) - F_a(r_1) \right)$$

$$= \lim_{a \to \infty} \int_{r_1}^{r_2} f_a(t) \, dt.$$

Qed

Theorem 3 (Very important)

Inversion formula for densities.

If X has a chf $\Phi_X(\cdot)$ s.t. $\int_{-\infty}^{\infty} |\Phi_X(t)|dt < \infty$

then X has a uniformly continuous density $f_X(\cdot)$ given by:

$$\boxed{f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\,\Phi_X(t)\,dt}$$

Proof: Recall that we have constructed

$$f_a(t) = c\int_{-\infty}^{\infty} e^{-itv}\,\Phi_X(v)\,f_W(av)\,dv$$

Now:

$$\left|e^{-itv}\,\Phi_X(v)\,f_W(av)\right| \le 1\cdot|\Phi_X(v)|\cdot C$$

By the hyp. on W

Then, by the DCT and ⊛ we have

$$f_a(t) \xrightarrow[a \to 0]{} c\, f_W(0) \int_{-\infty}^{\infty} e^{-itv}\, \Phi_X(v)\, dv$$

If we specify $U = W \sim N(0,1)$ we obtain $c f_W(0) = \frac{1}{2\pi}$

Thus, we showed that

$$f_a(t) \xrightarrow[a \to 0]{} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itv}\, \Phi_X(v)\, dv \equiv f(t).$$

- We now need to show that:
  1) f is uniformly continuous
  2) f is indeed the density of F_X.

See page 347 for 1)

For 2) first recall that we have proved

$$F_X(r_2) - F_X(r_1) = \lim_{a \to 0} \int_{r_1}^{r_2} f_a(t)\, dt.$$

Bonus exercise

If we show that $f_a \xrightarrow[a \to 0]{} f$ <u>uniformly</u>, then we can permute $\lim$ and $\int$ to obtain the desired result $F_X(r_2) - F_X(r_1) = \int_{r_1}^{r_2} f(t)\, dt$.

Qed

- Review of the $\boxed{\text{convolution}}$ theorem

1) Let ~~X~~ and Y be two independent r.v.s on $(\Omega, \mathcal{A}, P)$.
(We do <u>not</u> assume that they have densities). Let F_X and F_Y be the cdf's of X and Y, respectively.

$$F_{X+Y}(z) = P(X+Y \le z)$$

$$= \iint_{\{(x,y)\in\mathbb{R}^2 \mid x+y\le z\}} dF_X(x)\, dF_Y(y)$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} dF_Y(y) \right) dF_X(x)$$

$$= \int_{-\infty}^{\infty} F_Y(z-x)\, dF_X(x).$$

Given any two functions F_X and F_Y, the function F_{X+Y} given by:

$$F_{X+Y}(z) \equiv \int_{-\infty}^{\infty} F_Y(z-x)\, dF_X(x)$$

$$\equiv (F_X * F_Y)(z) \text{ is called}$$

the convolution of F_X and F_Y.

2) If now we further assume that Y has density f_Y with respect to the Lebesgue measure, then $X+Y$ will have a density wrt to the Lebesgue measure. (NOTE that we do not need to assume that X also has a density).

For this, notice that

$$F_{X+Y}(z) \underset{\substack{\uparrow \\ \text{as before}}}{=} \int_{-\infty}^{\infty} F_Y(z-x)\, dF_X(x)$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} f_Y(y)\, dy \right) dF_X(x)$$

$$\underset{\substack{\uparrow \\ v=y+x}}{=} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} f_Y(v-x)\, dv \right) dF_X(x)$$

$$\underset{\substack{\uparrow \\ v=y}}{=} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} f_Y(y-x)\, dy \right) dF_X(x)$$

$$= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} f_Y(y-x)\, dF_X(x) \right) dy$$

Hence, $F_{X+Y}(z)$ has density

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_Y(z-x)\, dF_X(x) \qquad (1)$$

3) If, in addition, $X \sim f_X(x)$, then (1) can be further simplified to:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_Y(z-x)\, f_X(x)\, dx$$

$$\equiv \left(f_Y * f_X\right)(z)$$

The convolution of the two densities.

<u>Moral</u>: a) For the sum of two independent rvs to have a density it is <u>enough</u> that one rv has density.

b) If they both have densities, f_{X+Y} has a "nicer" form.

- Now, by the definition of the chf, for a given F (the cdf of a rv X) we can compute (uniquely) its chf via

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x).$$

- We shall establish that the converse also holds: given a chf Φ, we can always find a df F that is uniquely determined by Φ

- We do so by establishing a formula that links Φ to F ~~in~~ an unique way. This will also provide an <u>inversion</u> formula. We accomplish that by giving a <u>constructive</u> proof, that is we will give an explicit link between Φ and F.

Strategy:

-) It is also nicer to work with rvs that have densities. However, not any arbitrary rv X has a density.

But: Via the convolution example, we know that the sum of 2 indep. rvs, one of which has a density, has a density itself.

Let X be an arbitrary r.v. with df F and let U be a r.v. with continuous density f_U, ~~then~~ independent of X.

Consider a perturbed version of X:

$$X_a \equiv X + aU.$$

As we have seen, X_a always has a density, f_a.

Moreovere, by Slutsky's Lemma:

$$X_a \xrightarrow{d} X \quad \text{as } a \to 0, \text{ or}$$

$$\lim_{a \to 0} F_a(\cdot) = F_X(\cdot), \quad \text{for any continuity point of } F_X \equiv F.$$

Hence: Instead of working with

$X \sim F$ directly

we work with

$X_a \sim f_a$, since $F_a \xrightarrow[a \to 0]{} F$.

Since we will eventually take a limit over $a \to 0$, U can be chosen in any convenient way. U is just a technical tool, and it will ALMOST dissappear from the final results.

Lecture 30

- Convergence in distribution via characteristic functions (Cramér- Levy).
- The Cramér - Wold device.

Theorem (Cramér- Levy).

1) **If** $\Phi_n \xrightarrow[n \to \infty]{} \Phi$, where $\{\Phi_n\}_n$ is a sequence of functions and Φ is a function continuous at 0,

then $F_n \xrightarrow[n \to \infty]{} F$.

. **Moreover**, Φ is the ch f associated with F.

2) If $F_n \longrightarrow F$, then $\Phi_n \longrightarrow \Phi$, uniformly over any finite interval $|t| \leq T < \infty$.

We have $\{X_n\}$ with df F_n

<u>Proof</u> We only prove 2).

- Recall that $F_n \to F$ is a notation for $X_n \xrightarrow{d} X$, for appropriately defined r.v's X_n and X.

- By Skorokhod's construction we can replace X_n and X by Y_n and Y s.t.

$$Y_n \xrightarrow{a.s.} Y \iff Y_n - Y \xrightarrow{a.s.} 0.$$

- By the "continuity" theorem:

$$e^{it(Y_n - Y)} \xrightarrow[n\to\infty]{a.s.} 1.$$

- Note also that $|e^{it(Y_n-Y)}| = 1$.

- Then, for any $|t| \le T$:

$$|\Phi_n(t) - \Phi(t)| \le \int |e^{itY_n} - e^{itY}| \, dP$$

$$= \int \underbrace{|e^{itY}|}_{=1} \cdot |e^{it(Y_n-Y)} - 1| \, dP$$

$$\le \int \sup_{|t| \le T} |e^{it(Y_n-Y)} - 1| \, dP$$

$$\xrightarrow[\text{By the DCT}]{} 0 \quad \left(\text{since } |e^{it(Y_n-Y)} - 1| \le 2 \text{ for all } |t| \le T.\right.$$

The characteristic function of $\underset{\sim}{X} \equiv (X_1, \ldots, X_K)$.

- Let $X_1, \ldots, X_K$ be rvs on $(\Omega, \mathcal{A}, P)$.

Let $\underset{\sim}{X} \equiv (X_1, \ldots, X_K) : \Omega \to \mathbb{R}^K$, and let $P_{\underset{\sim}{X}}$ be the induced measure.

$$(*) \quad \Phi_{\underset{\sim}{X}}(\underset{\sim}{t}) \equiv E\, e^{i \underset{\sim}{t}' \underset{\sim}{X}} \equiv E\, e^{i[t_1 X_1 + \ldots + t_K X_K]}, \quad \text{for } \underset{\sim}{t} \in \mathbb{R}^K, \; \underset{\sim}{t} \equiv (t_1, \ldots, t_K).$$

<u>Theorem</u>

$\Phi_{\underset{\sim}{X}}$ determines uniquely $P_{\underset{\sim}{X}}$. (See page 351.)

<u>Remark</u>.

Let $Y = \underset{\sim}{\lambda}' \underset{\sim}{X}$, for $\underset{\sim}{\lambda} \in \mathbb{R}^K$.

<u>Then</u>:

$$(*) \quad \Phi_Y(t) = E\, e^{i[\underbrace{t\lambda_1}_{t_1} X_1 + \ldots + \underbrace{t\lambda_K}_{t_K} X_K]}; \quad t \in \mathbb{R}, \; \lambda_j \in \mathbb{R}, \; 1 \le j \le K.$$

From (*) and (**):

knowing $\Phi_{\tilde{X}}(\underset{\sim}{t})$ for all $\underset{\sim}{t} \in \mathbb{R}^k$

(that is, knowing $P_{\tilde{X}}$) is the same

with knowing

$$\Phi_{\underset{\sim}{\lambda}'\underset{\sim}{X}}(t) \text{ for } \underline{\underline{\text{all}}} \quad \begin{matrix} t \in \mathbb{R} \\ \underset{\sim}{\lambda} \in \mathbb{R}^k \end{matrix} \; ; \; |\underset{\sim}{\lambda}| = 1.$$

(that is, knowing $P_{\underset{\sim}{\lambda}'\underset{\sim}{X}}$).

<u>Theorem 3</u> (VERY USEFUL for conv. in distribution in high dimensions)

Let $\underset{\sim}{X}_n \equiv (X_{n1}, \dots, X_{nk})'$.

<u>If</u> $\Phi_{\underset{\sim}{\lambda}'\underset{\sim}{X}_n}(t) \longrightarrow \Phi_{\lambda'\underset{\sim}{X}}(t)$ for all $t \in \mathbb{R}$ and each $\underset{\sim}{\lambda} \in \mathbb{R}^k$, with $|\underset{\sim}{\lambda}| = 1$, where $|\underset{\sim}{\lambda}|^2$ is the Euclidean norm.

<u>then</u> $\underset{\sim}{X}_n \xrightarrow{d} \underset{\sim}{X}$.

(This result is called the Cramér – Wold device).

__Proof__:

$$\Phi(t) = \int_{-\infty}^{\infty} e^{itx} \, dF(x).$$

Recall that:

$$\sin y = \sum_{n=0}^{\infty} (-1)^n \frac{y^{2n+1}}{(2n+1)!}, \quad y \in \mathbb{R}.$$

$$\cos y = \sum_{n=0}^{\infty} (-1)^n \frac{y^{2n}}{(2n)!}; \quad y \in \mathbb{R}.$$

Also:

$$e^{itx} = \cos tx + i \sin tx.$$

__Then__:

$$e^{itx} = \sum_{k=0}^{m-1} \frac{(itx)^k}{k!} + \frac{(itx)^m}{m!}\left[\cos \theta_1 tx + i \sin \theta_2 tx\right],$$

for $0 < |\theta_1| \vee |\theta_2| < 1$.

$$= \sum_{k=0}^{m} \frac{(itx)^k}{k!} + \frac{(itx)^m}{m!}\left[\cos(\theta_1 tx) + i \sin(\theta_2 tx) - 1\right]$$

Remark

Please note that, by the Cramer-Levy theorem, it is enough to show $\underset{\sim}{\lambda}' \underset{\sim}{X}_n \xrightarrow{d} \lambda' \underset{\sim}{X}$, no matter what method is used for this.

Proposition (Moment expansions of ch.fs).

Suppose $E|X|^m < \infty$ for some $m \geq 0$.

Then:

$$\frac{\left| \Phi(t) - \sum_{K=0}^{m} \frac{(it)^K}{K!} EX^K \right|}{|t|^m} \longrightarrow 0, \text{ as } t \to 0.$$

$$\Phi(t) = E e^{itX}$$

$$= E\left\{\sum_{K=0}^{m} \frac{(it)^K}{K!} \cdot X^K + \frac{(it)^m}{m!} X^m \left[\cos(\theta_1 t X) + i \sin(\theta_2 t X) - 1\right]\right\}$$

$$= \sum_{K=0}^{m} \frac{(it)^K}{K!} E X^K + \frac{(it)^m}{m!} E X^m \left[\cos(\theta_1 t X) + i \sin(\theta_2 t X) - 1\right]$$

$$\left|\Phi(t) - \sum_{K=0}^{m} \frac{(it)^K}{K!} E X^K\right| = \left|\frac{(it)^m}{m!}\right| \cdot \left|E X^m [\cdots\cdots]\right| =$$

$$\Rightarrow \frac{\left|\Phi(t) - \sum_{K=0}^{m} \frac{(it)^K}{K!} E X^K\right|}{|t|^m} = \frac{1}{m!} \left| \cdots \right|$$

It thus suffices to show that:

$$\lim_{t \to 0} E \left| X^m [\cos(\theta_1 t X) + i \sin(\theta_2 t X) - 1] \right| = 0,$$ which holds by our hyp. and the DCT.

Qed.

Corollary.

Assume X has mean 0 and variance σ^2.

The previous theorem states that:

$$\Phi(t) - \left[1 - \frac{\sigma^2 t^2}{2}\right] = o(t^2) \text{ as } t \to 0$$

Lecture 31

Central limit theorems via cfhs

<u>Theorem 1</u> (Classical CLT)

Let $X_{n1}, \ldots, X_{nn}$ be i.i.d. $F(\mu, \sigma^2)$ with $\mu < \infty$, $\sigma^2 < \infty$. Let $T_n \equiv X_{n1} + \ldots + X_{nn}$.

Let $\overline{X}_n \equiv \frac{T_n}{n}$.

<u>Then</u>

$$\sqrt{n}\,(\overline{X}_n - \mu) = \frac{1}{\sqrt{n}}\,(T_n - n\mu)$$

$$= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_{nk} - \mu) \xrightarrow[n \to \infty]{d} N(0, \sigma^2).$$

Proof

$$\Phi_{\sqrt{n}(\overline{X}_n-\mu)}(t) \underset{\substack{\uparrow \\ \text{indep}}}{=} \prod_{k=1}^{n} \Phi_{\left(\frac{X_{nk}-\mu}{\sqrt{n}}\right)}(t)$$

$$\underset{\substack{\text{identically} \\ \text{distr.} \\ +\text{change of var}}}{=} \left[\Phi_{X_{nk}-\mu}\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

$$\underset{\uparrow}{=} \left[1-\frac{\sigma^2}{2}\left(\frac{t}{\sqrt{n}}\right)^2+\left(\frac{t}{\sqrt{n}}\right)^2\cdot r\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

Lemma 13.4.2 + Ineq. 13.4.1
with the rest denoted
by $r(t)$; $r(t)\to 0$
as $t\to 0$

(Or Ineq 13.4.3)

$$=\left[1-\frac{\sigma^2 t^2}{2n}+\frac{t^2}{n}\cdot r\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

- Recall that

$$\lim_{b_n\to 0}\left(1+b_n\right)^{\frac{1}{b_n}}=e,$$

Let $b_n = -\frac{\sigma^2 t^2}{2n} + \frac{t^2}{n} \cdot r\left(\frac{t}{\sqrt{n}}\right)$.

Then:

$$(1+b_n)^n = \left[(1+b_n)^{\frac{1}{b_n}}\right]^{b_n \cdot n}$$

Thus $\lim_{n\to\infty} (1+b_n)^n = e^{\lim_{n\to\infty} b_n \cdot n}$; if the lim of the exponent exists.

$$\lim_{n\to\infty} n \cdot b_n = \lim_{n\to\infty} \left[-\frac{\sigma^2 t^2}{2} + t^2 \cdot r\left(\frac{t}{\sqrt{n}}\right)\right] = $$
$$= -\frac{\sigma^2 t^2}{2}.$$

Thus: $\Phi_{\sqrt{n}(\bar{X}_n - \mu)}(t) \to e^{-\frac{\sigma^2 t^2}{2}} = \Phi_{N(0,\sigma^2)}(t)$

$\Rightarrow$ By Th. 13.3.1 (Cramer-Levy) and uniqueness we have $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0,\sigma^2)$

Qed.

Theorem 2 (Classical Poisson limit theorem; PLT).

For each $n \geq 1$, let

$X_{n1}, \dots, X_{nk}, \dots X_{nn}$ be independent Bernoulli (λ_{nk}),

that is, $X_{nk} \begin{pmatrix} 0 & 1 \\ 1-\lambda_{nk} & \lambda_{nk} \end{pmatrix}$.

Assume that:

$$\begin{cases} \bullet \ \lambda_n \equiv \sum_{k=1}^{n} \lambda_{nk} \longrightarrow \lambda \\ \bullet \ \sum_{k=1}^{n} \lambda_{nk}^2 \longrightarrow 0. \end{cases} \qquad (1)$$

Then:

$$T_n \equiv X_{n1} + \dots + X_{nn} \xrightarrow{d} \text{Poisson}(\lambda).$$

Remark: 1) If $\lambda_{n1} = \dots = \lambda_{nn} = \dfrac{\lambda_n}{n}$ and $\lambda_n \longrightarrow \lambda$ then (1) holds.

2) If $X_1, \dots, X_n \sim$ Bernoulli(p) (that is, iid), then $T_n = \text{Bin}(np, np(1-p))$ and we know that $T_n \longrightarrow \text{Poisson}(\lambda)$, if $np \longrightarrow \lambda$.

Proof

We have $\Phi_{X_{nk}}(t) = 1 + \lambda_{nk}(e^{it} - 1)$.

$$\Phi_{T_n}(t) = \prod_{k=1}^{n} \Phi_{X_{nk}}(t)$$

$$= \prod_{k=1}^{n} [1 + \lambda_{nk}(e^{it} - 1)] \longrightarrow e^{\lambda(e^{it}-1)} = \Phi_{Pois(\lambda)}(t)$$

Lemma 4.3

$\overset{\text{Cramér-Levy}}{\Longrightarrow}$ $T_n \xrightarrow{d} \text{Poisson}(\lambda)$.

q.e.d.

Reading exercise: Exercise 1.4.

Theorem 3 (Classical multivariate CLT).

Let d be a fixed number.

Let $X_1, \dots, X_n$ be i.i.d (μ, Σ) vectors,

where $X_j \equiv (X_{1j}, \dots, X_{dj})$; $1 \le j \le n$;

$\mu \in \mathbb{R}^d$ and Σ is a $d \times d$ matrix.

Then $$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_j - \mu) \xrightarrow[n \to \infty]{d} N_d(0, \Sigma).$$

(Here $0 \in \mathbb{R}^d$).

Proof. Let $\lambda \in \mathbb{R}^d$ be arbitrary.

Then: $Y_j \equiv \lambda'(X_j - \mu) \cong (0, \lambda'\Sigma\lambda)$ are iid vectors, $1 \le j \le n$.

- Applying the univariate CLT to $\{Y_j\}_{1 \le j \le n}$ we obtain:

$$\sqrt{n}\, \overline{Y}_n \xrightarrow{d} N(0, \lambda'\Sigma\lambda). \qquad (1)$$

- Notice now that:

$$\sqrt{n}\,\bar{Y}_n = \sqrt{n}\cdot\frac{1}{n}\sum_{j=1}^{n} Y_j$$

$$= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\lambda'(X_j-\mu)$$

$$= \lambda'\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j-\mu) \equiv \lambda' Z_n .$$

- Therefore, by the Cramér-Levy theorem and ①:

$$② \quad \Phi_{\lambda' Z_n}(t) \xrightarrow[n\to\infty]{} e^{-\frac{\lambda'\Sigma\lambda t^2}{2}} \equiv \Phi_{\lambda' Z}(t), \quad \text{with}$$

$Z \cong N_d(0, \Sigma)$; for all $\lambda \in \mathbb{R}^k$, $\forall t$.

- Thus, by the Cramér-Wold device, ② implies that $Z_n \xrightarrow{d} Z$.

q.e.d.

Lecture 32

Conditional Probability and Expectation

Definition 1 (Conditional Expectation.)

- Let $(\Omega, \mathcal{A}, P)$ denote a probability space.
- Let $\mathcal{D}$ denote a σ-field of $\mathcal{A}$.
- Let Y be a r.v. on $(\Omega, \mathcal{A}, P)$ for which $E|Y| < \infty$.
- The conditional expectation

$$E(Y/\mathcal{D})(\cdot)$$

is a $\mathcal{D}$-measurable function s.t.:

① $$\int_D E(Y/\mathcal{D})(\omega)\, dP(\omega) = \int_D Y(\omega)\, dP(\omega)$$

for all $D \in \mathcal{D}$.

Remark 1

Recall that a r.v. is

$$Y : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}).$$

Let X be another r.v. on $(\Omega, \mathcal{A}, P)$.
Then $\mathcal{F}(X) \equiv X^{-1}(\mathcal{B}) \subseteq \mathcal{A}$.
Thus, in the above definition we take $\mathcal{D} = \mathcal{F}(X)$ we have the following notation:

$$E(Y \mid \mathcal{F}(X))(\cdot) \underset{\substack{\uparrow \\ \text{Notation}}}{\equiv} E(Y \mid X)(\cdot).$$

NB1 We shall use this notation all the time.

NB2 The conditional expectation is a function (a r.v.) and NOT a number (as is the case with expectations).

<u>Remark 2</u> The conditional expectation is well defined. We show below that $E(Y|\mathcal{D})(\cdot)$ exists and is unique a.s.

- First let $Y \geq 0$.
- Define a measure on $\mathcal{D}$ by:

$$\nu(D) \equiv \int_D Y\,dP \quad \text{for all } D \in \mathcal{D}.$$

We have seen that this defines a measure.

- Since P is a measure on $\mathcal{A}$, $P_{|\mathcal{D}}$ (the restriction of P to $\mathcal{D}$) is also a measure on $\mathcal{D}$.

- Notice now that, for these two measures on $\mathcal{D}$, we have:

$$\nu \ll P_{|\mathcal{D}}$$

• Thus, by the R–N theorem,

there exists, uniquely a.s. $P_{|\mathcal{D}}$,

a $\mathcal{D}$-measurable function h such that

$$\nu(D) = \int_D h \, dP_{|\mathcal{D}} = \int_D h \, dP .$$

• The function h is $\mathcal{D}$-measurable and unique a.s. $P_{|\mathcal{D}}$ $\Big\} \Longrightarrow$

h is unique a.s. P.

Thus: we have proved the existence of h, that is a.s. P unique, such that:

$$\int_D h \, dP = \int_D Y \, dP, \text{ for all } D \in \mathcal{D}.$$

We denote h by $E(Y|\mathcal{D})$.

• For general Y, we use the general $Y = Y^+ - Y^-$.

Everett Pad & Paper Company
1-888-387-8723

Very important proposition.

Prop. 2.2.5

If Z is a r.v. on $(\Omega, \mathcal{A})$ that is $\mathcal{F}(X)$ measurable. Then $\exists$ a measurable g on $(\mathbb{R}, \mathcal{B})$ s.t. $Z = g(X)$.

A $\mathcal{F}(X)$ measurable function is just some function of X.

Remark 3 (important).

There exists a measurable function $\underline{g}$, on $(\mathbb{R}, \mathcal{B})$, such that:

- $E(Y|X)(\omega) = g(X(\omega))$ *

• Remark : One typically encounters the following notation : $E(Y|X=x)$. What does this mean in view of the above considerations ?

Definition 1'

We can define $g(x) = E(Y|X=x)$ to be the $\mathcal{B}$-measurable ~~function~~ function on $\mathbb{R}$ s.t.:

$$\int_B E(Y|X=x)\,dP_X(x) = \int_{X^{-1}(B)} Y(\omega)\,dP(\omega),$$

for all $B \in \mathcal{B}$.

• The considerations below justify this definition.

- By definition 1 of conditional expectation

$$E(Y/X) = E(Y/\underline{\underline{\mathcal{F}(X)}}) \text{ and is } \mathcal{F}(X) \text{ measurable.}$$

By Prop. 2.2.5 $\Rightarrow$ $\exists$ g measurable, on $(\mathbb{R}, \mathcal{B})$, such that:

$$h \equiv E(Y/X) = g(X). \quad (2).$$

- Let $D \in \mathcal{F}(X)$ be an arbitrary element. We know that $D = X^{-1}(B)$, $B \in \mathcal{B}$.

- By the Th. of the unconscious statistician:

$$\int_{X^{-1}(B)} g(X)(\omega)\, dP(\omega) = \int_B g(x)\, dP_X(x) \qquad (3)$$

- Go now back to the definition (1) page 1, for $D = X^{-1}(B)$ and $\mathcal{D} = \mathcal{F}(X)$. Using (2) and (1) we have

$$\int_{X^{-1}(B)} g(X)(\omega)\, dP(\omega) = \int_{X^{-1}(B)} Y(\omega)\, dP(\omega) \qquad (4)$$

Thus, by (3) and (4):

$$(5) \quad \int_{X^{-1}(B)} Y(\omega)\, dP(\omega) = \int_B g(x)\, dP_X(x), \quad \forall\, B \in \mathcal{B}.$$

Such g exists and is unique a.s. P_X by the above.

- We can now take (5) as Definition 1'.

Definition 2

Let $(\Omega, \mathcal{A}, P)$ be a prob. space.
Let $\mathcal{D} \subset \mathcal{A}$ be a sub σ-field.
For any $A \in \mathcal{A}$ define:

$$P(A \mid \mathcal{D}) = E(1_A \mid \mathcal{D}).$$

- **Remark** By definition ① we thus have that $\boxed{P(A \mid \mathcal{D}) \text{ is a } \mathcal{D} \text{ measurable function}}$

such that; for any $D \in \mathcal{D}$ we have:

$$\int_D P(A \mid \mathcal{D}) = \int_D 1_A \, dP \qquad \text{or:}$$

$$\boxed{P(A \cap D) = \int_D P(A \mid \mathcal{D}) \, dP}, \text{ for all } D \in \mathcal{D}.$$

- As before, the conditional probability exists and is unique a.s. P.

Corollary (of the definition...).

For $\mathcal{D} = \mathcal{F}(X)$, we also denote

$$P(A \mid X) \equiv P(A \mid \mathcal{F}(X)).$$

As before:

$$P(A \mid X)(\omega) = g(X(\omega)), \text{ where}$$

$g(x) = P(A \mid X = x)$ is a $\mathcal{B}$-measurable function satisfying

$$P(A \cap X^{-1}(B)) = \int_B P(A \mid X = x)\, dP_X(x),$$

for all $B \in \mathcal{B}$.

N.B.1: $P(A \mid X)$ is a measurable function, for any fixed and given A.

NB 2 :

(1) $0 \leq P(A \mid \mathcal{D}) \leq 1$ a.s. P

(2) $P(\sum_{i=1}^{\infty} A_i \mid \mathcal{D}) = \sum_{i=1}^{\infty} P(A_i \mid \mathcal{D})$ a.s. P

(3) $P(\emptyset \mid \mathcal{D}) = 0$, a.s. P

(4) $A_1 \subset A_2 \Rightarrow P(A_1 \mid \mathcal{D}) \leq P(A_2 \mid \mathcal{D})$ a.s. P

(So, the measurable function $P(A \mid \mathcal{D})$ has properties resembling a probability distribution.)

<u>Reading assignment</u> : Pages 160-161.
Pages $\underbrace{151-154, 157}$
see below for highlights.

Some useful definitions and results on independence.

- Let $(\Omega, \mathcal{K}, P)$ be a given probability space.

1) **Independence of σ-fields.**

Let $\mathcal{A}_1, \dots, \mathcal{A}_n$ be sub-σ-fields of $\mathcal{K}$.
We say that $\mathcal{A}_1, \dots, \mathcal{A}_n$ are independent if

$$P(A_1 \cap \dots \cap A_n) = \prod_{i=1}^{n} P(A_i)$$

for all $A_i \in \mathcal{A}_i$; $1 \le i \le n$.

2) **Independence of random variables**

Let $X_1, \dots, X_n$ be random variables.
For each X_i consider the σ-field

$\mathcal{F}(X_i) \equiv X_i^{-1}(\mathcal{B})$.

$X_1, \dots, X_n$ are called independent if $\mathcal{F}(X_1), \dots, \mathcal{F}(X_n)$ are independent.

3) <u>Independence of events</u>

The events $A_1, \dots, A_n$ are called independent if the σ-fields $\sigma[A_1], \dots, \sigma[A_n]$ are independent.

<u>Required</u> reading exercises (These exercises show that the complicated definitions 1) – 3) reduce to familiar things related to independence):

- Exercise 8.1.1 page 152.
- Theorem 8.1.1 page 152.
- Theorem 8.1.2 page 153
- Corollary 8.1 page 153.
- Criteria for independence : page 154.
- Chapter 8.3 page 157.

$\Phi_{\sum X_i}(t) = \prod_{i=1}^{n} \phi_{X_i}(t)$ for $X_i \cdots$ indep

Know the results.

Lecture 33

Properties of conditional expectation

(See Theorem 8.4.1 page 163 for the full statement. Here we give the proof of selected properties).

Proof

(1) Linearity

$$E(a \cdot X + bY \mid \mathcal{D}) = a\,E(X|\mathcal{D}) + b\,E(Y|\mathcal{D}) \quad \text{a.s. } P.$$

$$\int_D \{a\,E(X|\mathcal{D}) + b\,E(Y|\mathcal{D})\}\,dP$$

$$= a\int_D E(X|\mathcal{D})\,dP + b\int_D E(Y|\mathcal{D})\,dP$$

$$\underset{\text{definition}}{=} a\int_D X\,dP + b\int_D Y\,dP \underset{\substack{\text{lin. of integral} \\ \text{+ definition}}}{=} \int_D E(aX+bY \mid \mathcal{D})\,dP.$$

Everett Pad & Paper Company
1-888-387-8723

(2) $EY = E(E(Y|\mathcal{D}))$ //

$$EY = \int_{\Omega} Y\,dP \underset{\text{definition}}{=} \int_{\Omega} E(Y|\mathcal{D})\,dP$$

$$= E(E(Y|\mathcal{D})).$$

(3) $X \le Y$ ~~a~~ a.s.P $\Rightarrow E(X|\mathcal{D}) \le E(Y|\mathcal{D})$ a.s.P

(4) MCT

If $0 \le Y_n \nearrow Y$ a.s. P then

$E(Y_n|\mathcal{D}) \uparrow E(Y|\mathcal{D})$ a.s.P //

Since $Y_n \uparrow Y \Rightarrow$

$$Y_n \le Y_{n+1} \le Y \quad \text{a.s.P}$$

$$\Downarrow (3)$$

$$0 \le E(Y_n|\mathcal{D}) \le E(Y_{n+1}|\mathcal{D}) \le E(Y|\mathcal{D}) \quad \text{a.s. P}$$

$$\Downarrow$$

$\{E(Y_n|\mathcal{D})\}_n$ has a limit.

We show now that it is the right limit:

$$\int_D \underline{\underline{\lim_n E(Y_n|\mathcal{D})}}\, dP \underset{\substack{\uparrow \\ \text{MCT for } \{E(Y_n|\mathcal{D})\}_n}}{=} \lim_n \int_D E(Y_n|\mathcal{D})\, dP$$

$$\underset{\substack{\uparrow \\ \text{Definition}}}{=} \lim_n \int_D Y_n\, dP$$

$$\underset{\substack{\uparrow \\ \text{MCT for } \{Y_n\}_n}}{=} \int_D Y\, dP$$

$$= \int_D \underline{\underline{E(Y|\mathcal{D})}}\, dP \quad \overset{\text{Exercise 3.2.}}{\Longrightarrow}$$

$$\lim_n E(Y_n|\mathcal{D}) = E(Y|\mathcal{D}) \quad \text{a.s. } P$$

Show for indicators
Show for elementary (simple funct^{n})
Show for +ve. funct^{n}s using MCT
Show for a general function

(20) If Y is $\mathcal{D}$ - measurable and $XY \in \mathcal{L}_1(P)$, then

$$E(XY|\mathcal{D}) = Y E(X|\mathcal{D}) \text{ a.s.}$$

Proof It is enough to show that (20) holds for indicators. Then we can apply the MCT for the general case.

Let $Y = 1_{D^*}$, for some fixed D^*, associated with Y, in $\mathcal{D}$.

Let D be an arbitrary set in $\mathcal{D}$.

Then:

$$\int_D E(XY|\mathcal{D}) = \int_D E(X \cdot 1_{D^*} | \mathcal{D})$$

$$\underset{\substack{\uparrow \\ \text{def. of} \\ \text{cond. expectation}}}{=} \int_D X \cdot 1_{D^*} \, dP = \int_{\substack{DD^* \\ \in \mathcal{D}}} X \, dP$$

$$\underset{\substack{\uparrow \\ \text{def of} \\ \text{cond. expectation}}}{=} \int_{DD^*} E(X|\mathcal{D}) \, dP$$

$$= \int_D 1_{D^*} E(X|\mathcal{D})\, dP.$$

<u>Thus</u>: $\int_D E(X \cdot 1_{D^*}|\mathcal{D}) = \int_D 1_{D^*} E(X|\mathcal{D})\, dP. \Longrightarrow$

Exercise 3.2.2

$$E(X \cdot 1_{D^*}|\mathcal{D}) = 1_{D^*} E(X|\mathcal{D}) \quad \text{a.s.}$$

(23) If $\mathcal{F}(Y, X_1)$ is independent of $\mathcal{F}(X_2)$, then
$E(Y|X_1, X_2) = E(Y|X_1)$ a.s.

<u>Proof</u>: Recall that:

$$E(Y|X_1, X_2) \equiv E(Y|\mathcal{F}(X_1, X_2))$$

$$E(Y|X_1) \equiv E(Y|\mathcal{F}(X_1)).$$

- <u>idea</u>: To show (23) it is enough to show that

$$\int_D E(Y|\mathcal{F}(X_1, X_2))\, dP = \int_D E(Y|\mathcal{F}(X_1))\, dP,$$

for any $D \in \mathcal{F}(X_1, X_2)$, then we apple 3.2.2

- Recall that

$$\mathcal{F}(X_1, X_2) = \sigma[D], \quad \text{where}$$

$$(1): \quad D \equiv D_1 \cap D_2 \equiv X_1^{-1}(B_1) \cap X_2^{-1}(B_2), \qquad B_1, B_2 \in \mathcal{B}$$

- Also recall that given a generic r.v. Z ($\mathcal{D}$ measurable) we can define a measure on $\mathcal{D}$.

$$\nu(D) \equiv \int_D Y \, dP, \quad \text{for } D \in \mathcal{D}.$$

- Now, consider the collection of sets D given by (1) above.

- Define $\nu_1(D) \equiv \int_D E(Y \mid \mathcal{F}(X_1, X_2)) \, dP$

$$\nu_2(D) \equiv \int_D E(Y \mid \mathcal{F}(X_1)) \, dP.$$

- If we show that $\nu_1(D) = \nu_2(D)$, for all D in this collection, and if the collection forms a π-system, then (by the Caratheodory theorem), $\nu_1(D) = \nu_2(D)$ for all $D \in \mathcal{F}(X_1, X_2)$, which is what we need to show.

It thus suffices to show that
$\nu_1(D) = \nu_2(D)$, for all D that generate
$\sigma[D] \equiv \mathcal{F}(X_1, X_2)$.

$$\nu_1(D) = \int_D E(Y|X_1,X_2)\,dP \underset{\uparrow}{=} \int_D Y\,dP$$

def of cond. expectation

$$= \int_{D_1\cap D_2} Y\,dP = \int 1_{D_1\cap D_2} \cdot Y\,dP$$

$$= \int \underset{X_2^{-1}(B_2)}{1_{D_2}} \cdot \Big(\underset{\underset{X_1^{-1}(B_1)}{\parallel}}{1_{D_1}} \cdot Y\Big)\,dP \underset{\uparrow}{=}$$

The independence hypothesis

$$= \int 1_{D_2}\,dP \cdot \int 1_{D_1} Y\,dP = \int 1_{D_2}\,dP \cdot \int_{D_1} Y\,dP$$

$$\underset{\uparrow}{=} \int 1_{D_2}\,dP \cdot \int_{D_1} E(Y/X_1)\,dP \underset{\uparrow}{=} \int 1_{D_2} \cdot 1_{D_1} \cdot E(Y/X_1)\,dP$$

def. of cond. expectation for $E(Y/X_1)$

Indep. again

$$= \int_{D_1 D_2 = D} E(Y/X_1)\,dP = \nu_2(D).$$

qed

Lecture 33' Not Req

Regular conditional distributions (Optional reading).

Definition 1 (Reg. cond. distr.).

Let $Z: (\Omega, \mathcal{A}) \to (M, \mathcal{S})$. Let $\mathcal{D}$ be a sub σ-field of $\mathcal{A}$. Let $\tilde{\mathcal{A}} \equiv Z^{-1}(\mathcal{S})$.

Then $P_Z(S|\mathcal{D})(\omega)$ will be called a regular conditional distribution for Z given $\mathcal{D}$ if:

(1) For each fixed $S \in \mathcal{S}$ the function $P_Z(S|\mathcal{D})(\cdot)$ is a version of $P(Z \in S|\mathcal{D})(\cdot)$ on Ω. (and so it satisfies the def. of cond. prob.)

(2) For each fixed $\omega \in \Omega$ the set function $P_Z(\cdot|\mathcal{D})(\omega)$ is a probability distribution on $(M, \mathcal{S})$.

Notation : $P_Z^{\omega}(\cdot \mid \mathcal{D}) \equiv P_Z(\cdot \mid \mathcal{D})(\omega)$.

Theorem If $Z : (\Omega, \mathcal{A}) \to (M, \mathcal{T})$ and $(M, \mathcal{T})$ is a Borel space, then a regular conditional probability $P_Z(S \mid \mathcal{D})(\omega)$ exists.

Consequence : Elementary conditional densities (rigorous derivation!)

Assume $P((X,Y) \in B_2) = \iint_{B_2} f(x,y)\, dx\, dy$, for all $B_2 \in \mathcal{B}_2$, and $f \geq 0$ measurable; $f(x,y) \equiv$ the joint density.

- $\dfrac{dP_X(x)}{d\lambda} = f_X(x) \equiv \int_{\mathbb{R}} f(x,y)\, dy$. = the marginal density of X.

- Define : $g(y \mid x) \equiv \begin{cases} \dfrac{f(x,y)}{f_X(x)}, & \text{for all } x \text{ with } f_X(x) \neq 0 \\ \text{some arbitrary density } f_0(y), & \text{for } x \text{ with } f_X(x) = 0 \end{cases}$

$g(y \mid x) =$ the conditional density of Y given $X = x$.

<u>Define</u>

$$P(Y \in A \mid X = x) \equiv \int_A g(y|x)\, dy \;; \text{ For } A \in \mathcal{B}. \quad (10)$$

<u>Then</u>:

1) $P(Y \in A \mid X = x)$, given modified appropriately on the set where $f_X(x) = 0$, is a regular conditional distrib.

Moreover,

2) If $E|h(Y)| < \infty$, then

$$E(h(Y) \mid X = x) = \int_{-\infty}^{\infty} h(y)\, g(y|x)\, dy \;, \quad \text{a.s. } P_X.$$

<u>Proof</u>: In order to show that $P(Y \in A \mid X = x)$ is a reg. cond. distr. we first need to show that (10) does indeed define a cond distribution as in Def 8.4.2. page 159.

We know that, by that definition, that there exists a version such that:

$$(11) \quad \int_B P(Y \in A \mid X = x)\, dP_X(x) = P(A \cap X^{-1}(B)) = P([Y \in A] \cap X^{-1}(B))$$

for all $B \in \mathcal{B}$.

So, (11) is always true by definition.
We show that, in addition, (10) holds.

$$\underline{\underline{\int_B \left(\int_A g(y|x)\,dy\right) dP_X(x)}}$$

$$= \int_B \left[\int_A g(y|x)\,dy\right] f_X(x)\,dx$$

$$= \int_{B\cap S} \left[\int_A \frac{f(x,y)}{f_X(x)}\,dy\right] \cancel{f_X(x)}\,dx; \quad \text{on } S \equiv \{x \mid f_X(x)$$

$$= \iint_{B\cap A} f(x,y)\,dx\,dy = P((X,Y) \in B \times A)$$

$$\underline{= P((X\in B)\cap(Y\in A)) = P([Y\in A] \cap X^{-1}(B))}$$

$$\underset{\substack{\uparrow \\ \text{def of} \\ \text{cond. probab.}}}{=} \underline{\underline{\int_B P(Y\in A \mid X = x)\,dP_X(x)}}$$

Thus, indeed $P(Y \in A \mid X = x) \underset{\substack{\uparrow \\ \text{(given by this formula)}}}{=} \int_A g(y|x)\,dy$ a.s. P_X,

and so it is a version of what we needed.

<u>Moreover</u>, for any <u>fixed</u> $A \in \mathcal{B}$ we have:

$$(*) \quad \underbrace{\int_A g(y|x)\,dy}_{LHS} = 1_S(x) \cdot \int_A \frac{f(x,y)}{f_X(x)}\,dy + 1_{S^c}(x) \cdot \int_A f_0(y)\,dy$$

Thus, for any fixed x the LHS is given by only one of the terms to the right of the $=$ sign, and hence it is a prob. distribution.

- $(*)$ is really the defining quantity for a regular conditional probability distribution.

2) This formula shows that conditional expectations can be computed as expectations w.r.t conditional distr. functions. $\rightarrow$ See Th. 8.5.3 for a general statement.

- We show that, for any $B \in \mathcal{B}$ we have

$$\int_B E(h(Y) \mid X = x)\,dP_X = \int_B \left(\int_{-\infty}^{\infty} h(y) g(y|x)\,dy \right) dP_X$$

On the one hand:

$$\int_B \left(\int_R h(y) g(y|x)\, dy \right) dP_X(x) =$$

$$= \int_{B\cap S} \left[\int_R h(y) g(y|x)\, dy \right] f_X(x)\, dx$$

$$= \int_{B\cap S} \left[\int_R h(y) f(x,y)\, dy \right] dx$$

$$= \int_B \left(\int_R h(y) f(x,y)\, dy \right) dx$$

$$= \iint_{B\times R} h(y) f(x,y)\, dx\, dy$$

On the other hand:

$$\int_B E(h(Y) | X = x)\, dP_X(x) \underset{\uparrow}{=} \int_{X^{-1}(B)} h(Y)(\omega)\, dP(\omega)$$

By definition

Note that:

$$\int_B \left(\int_{\mathbb{R}} h(y)\, g(y|x)\, dy \right) dP_X(x) = \int_B \left[\int_{\mathbb{R}} h(y)\, f(x,y)\, dy \right] dx$$

$$= \int_B \int_{\mathbb{R}} h(y)\, f(x,y)\, dy\, dx = \int_B \int_{\mathbb{R}} h(y)\, dP_{X,Y}(x,y)$$

$$= \int_B E(h(Y) \mid X = x)\, dP_X(x)$$

qed.

Lecture 34

Conditional expectation as a projection operator

Not req

(Optional reading).

- Let
$$\mathcal{H} \equiv \mathcal{L}_2(P) \equiv \{X \text{ on } (\Omega, \mathcal{A}, P) \mid E X^2 < \infty\}.$$
- $\mathcal{H}$ is a Hilbert space with inner product
$$\langle X, Y \rangle \equiv E(XY).$$
- Let $\mathcal{D} \subset \mathcal{A}$. Define
$$\mathcal{H}_{\mathcal{D}} \equiv \{X \in L_2(P) : \mathcal{F}(X) \subset \mathcal{D}\} \subset \mathcal{H}.$$
(The space of all r.v.'s X that are $\mathcal{D}$ measurable)

- Define the <u>linear operator</u> $P_{\mathcal{D}}$, on $L_2(P)$, by:

$$P_{\mathcal{D}} Y \equiv E(Y|\mathcal{D}), \text{ for each } Y \in L_2(P).$$

- <u>$P_{\mathcal{D}}$ projects $\mathcal{H}$ onto $\mathcal{H}_{\mathcal{D}}$</u>.

<u>Proposition 1</u> (Properties of the projection operator).

1) $P_{\mathcal{D}}^2 = P_{\mathcal{D}}$ ($P_{\mathcal{D}}$ is idempotent ← this qualifies $P_{\mathcal{D}}$ to be a projection).

2) $I - P_{\mathcal{D}}$ is the projection onto $\mathcal{H}_{\mathcal{D}}^{\perp}$ (where I is the identity function and $\mathcal{H}_{\mathcal{D}}^{\perp}$ is the orthogonal complement of $\mathcal{H}_D$, that is:

$$\mathcal{H}_{\mathcal{D}} + \mathcal{H}_{\mathcal{D}}^{\perp} = \mathcal{H} \text{ and } \left.\begin{array}{l} \forall\, X \in \mathcal{H}_D \\ Y \in \mathcal{H}_D^{\perp} \end{array}\right\} \Rightarrow \langle X, Y \rangle = 0,$$

3) $P_{\mathcal{D}}$ is self adjoint: $P_{\mathcal{D}}^{*} = P_{\mathcal{D}}$, that is:

$$\langle P_{\mathcal{D}} Y, Z \rangle = \langle Y, P_{\mathcal{D}} Z \rangle, \text{ for all } Y \text{ and } Z \text{ in } \mathcal{H}.$$

4) a) $\| P_{\mathcal{D}} Y \| \leq \| Y \|$, for all Y.

b) $\| P_{\mathcal{D}} \| = 1$, if $\dim(\mathcal{H}_{\mathcal{D}}) \geq 1$, where

$$\| P_{\mathcal{D}} \| = \sup_{\substack{Y \in \mathcal{H} \\ Y \neq 0}} \left(\frac{\| P_{\mathcal{D}} Y \|}{\| Y \|} \right).$$

<u>Proof</u>:

1) Recall Th. 8.4.1 page 163.

Then: By (14), we have that $P_{\mathcal{D}}$ is linear.

$$\begin{aligned} P_{\mathcal{D}}^2 Y = P_{\mathcal{D}} P_{\mathcal{D}} Y &= P_{\mathcal{D}}(P_{\mathcal{D}} Y) \\ &= E(P_{\mathcal{D}} Y \mid \mathcal{D}) = E(E(Y \mid \mathcal{D}) \mid \mathcal{D}) \\ &\underset{\substack{\uparrow \\ (22) \text{ of Th. } 8.4.1}}{=} E(Y \mid \mathcal{D}) = P_{\mathcal{D}}(Y). \end{aligned}$$

2) First note that $I - P_{\mathcal{D}}$ is idempotent, hence it is a projection:

$$\begin{aligned} (I - P_{\mathcal{D}})^2 = I^2 - 2 I P_{\mathcal{D}} + P_{\mathcal{D}}^2 &= I - 2 P_{\mathcal{D}} + P_{\mathcal{D}}^2 \\ &\underset{\substack{\uparrow \\ P_{\mathcal{D}} = P_{\mathcal{D}}^2}}{=} I - P_{\mathcal{D}}^2 \underset{\substack{\uparrow \\ P_{\mathcal{D}}^2 = P_{\mathcal{D}}}}{=} I - P_{\mathcal{D}}. \end{aligned}$$

• To show that it is a projection <u>onto</u> $\mathcal{H}_D^\perp$, we need to show that:

$$\langle \underbrace{(I-P_{\mathcal{D}})Y}, Z \rangle = 0, \text{ for any } Y \in \mathcal{H}, \; Z \in \mathcal{H}_D.$$

$$\langle (I-P_{\mathcal{D}})Y, Z \rangle = \langle Y - P_{\mathcal{D}}Y, Z \rangle$$

$$\underset{\substack{\uparrow \\ \text{def. of} \\ \text{the inner product}}}{=} E((Y-P_{\mathcal{D}}Y)Z)$$

$$\underset{\substack{\uparrow \\ \text{def. of } P_{\mathcal{D}}Y}}{=} E[(Y - E(Y|\mathcal{D})) \cdot Z]$$

$$= E(YZ) - E(Z\,E(Y|\mathcal{D}))$$

$$\underset{\substack{\uparrow \\ Z \in \mathcal{H}_D, \\ \text{hence } Z \text{ is } \mathcal{D} \text{ measurable} \\ ((20) \text{ of Th. } 8.4.1)}}{=} E(YZ) - E(E(YZ|\mathcal{D}))$$

$$= E(YZ) - E(YZ) = 0 \quad //$$

This shows that the element $(I-P_{\mathcal{D}})Y \in \mathcal{H}_D^\perp$, as described in 2) at page 2.

3) $\langle P_{\mathcal{D}} Y, Z \rangle = E(\cancel{E}(Y|\mathcal{D}) \cdot \cancel{Z}) \underset{(8.4.1(20))}{=} E(E($

$= E(ZY)$

$\underset{(8.4.1,(20))}{=} E(Y E(Z|\mathcal{D})) = \langle Y, P_{\mathcal{D}} Z \rangle$

4) $\underline{\underline{\|P_{\mathcal{D}} Y\|^2}} \underset{\text{def of the } L_2 \text{ norm}}{=} E[(E(Y|\mathcal{D}))^2] \leq E[E(Y^2|\mathcal{D})]$

Jensen's Ineq. for cond. expect (Th. 8.4.1, (24))

$= E(Y^2) = \underline{\underline{\|Y\|^2}}.$

<u>Hence</u> $\|P_{\mathcal{D}} Y\| \leq \|Y\| \Rightarrow \dfrac{\|P_{\mathcal{D}} Y\|}{\|Y\|} \leq 1$, for all Y

$\Rightarrow \|P_{\mathcal{D}}\| \leq 1$, that is, $P_{\mathcal{D}}$ is a bounded operator

That is always valid!

- Furthermore, if we take $Y \neq 0$, $Y \in \mathcal{H}_{\mathcal{D}}$ (that is, if we further assume that Y is $\mathcal{D}$ measurable) we have $\|P_{\mathcal{D}} Y\|^2 = E(\underbrace{E(Y|\mathcal{D})}_{Y} \underbrace{E(Y|\mathcal{D})}_{Y}) = E($ $= \|Y\|$

$\Rightarrow \underline{\|P_{\mathcal{D}}\| = 1}$, (To conclude that, we must have Y's in $\mathcal{H}_{\mathcal{D}}$, that's why we need the assumption $\dim(\mathcal{H}_{\mathcal{D}}) \geq 1$).

Important Subject Matter

Projection on a Sum space by alternating projections : ACE

ACE = alternating conditional expectation.

- Let $Y, X_1, X_2, \ldots$ defined on $(\Omega, \mathcal{A}, P)$.
- The following spaces and projections are of interest:

$$\mathcal{H}_1 \equiv \mathcal{H}\mathcal{F}(X_1), \qquad P_1 Y \equiv E(Y|X_1)$$
$$\mathcal{H}_2 \equiv \mathcal{H}\mathcal{F}(X_2); \qquad P_2 Y \equiv E(Y|X_2),$$

where $Y \in \mathcal{H} \equiv L_2(P)$.

- Consider the <u>sum space</u>:

$$\mathcal{H}_+ \equiv \mathcal{H}_1 + \mathcal{H}_2 = \{g_1(X_1) + g_2(X_2) : g_i \text{ is measurable}; E g_i^2(X_i) < \infty\}$$

- Denote by P_+ the projection onto $\mathcal{H}_+$.

$$P_+ \equiv P_{\mathcal{H}_+}.$$

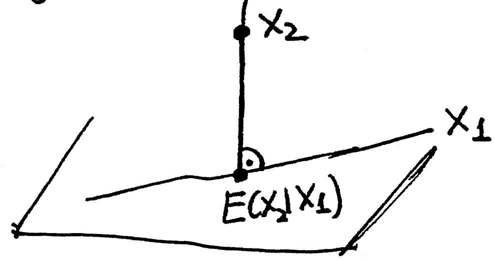
Can we add the regressions when the covariates are not independent.

If X_1 and X_2 are indep and we want a regression of Y on X_1 & X_2
We can $Y \sim X_1$ and $Y \sim X_2$ and then add the regression

• REMARK: X_1 and X_2 independent $\Longleftrightarrow$ $\mathcal{H}_1$ and $\mathcal{H}_2$ are orthogonal

Note that: $\mathcal{H}_1$ and $\mathcal{H}_2$ are orthogonal $\Leftrightarrow$

$$\langle X_2 - E(X_2|X_1), X_1 \rangle = 0$$



Now:

$$\langle X_2 - E(X_2/X_1), X_1 \rangle = \langle X_2, X_1 \rangle - \langle \underbrace{E(X_2|X_1)}_{\overset{\text{indep.}}{=} E(X_2)}, X_1 \rangle$$

$$= E(X_1 X_2) - EX_1 \cdot E(X_2) \underset{\text{indep}}{=} 0 \; //.$$

Proposition

If X_1 and X_2 are independent, then

$$P_+ Y = P_1 Y + P_2 Y, \text{ for all } Y \in \mathcal{H}.$$

NB1 If the hyp. fails the condition is no longer true (note that a projection is a linear operator in Y, not in $\mathcal{D}$!!).

NB2 Since $g_1(X_1)$ and $g_2(X_2)$ have finite mean we will assume in the proof that they are zero.

Proof

Recall that

$$P_+ \equiv P_{\mathcal{H}_+} \quad ; \quad \mathcal{H}_+ = \{ Z = g_1(X_1) + g_2(X_2) \}.$$

To show that $P_+ = P_1 + P_2$ ~~is~~ is to show that

$$\langle Y - E(Y|X_1) - E(Y|X_2), Z \rangle = 0, \text{ for all } Z \in \mathcal{H}_+$$

Then :

$$\langle Y - E(Y|X_1) - E(Y|X_2), g_1(X_1) + g_2(X_2) \rangle =$$

$$= \langle Y - E(Y|X_1), g_1(X_1) \rangle - \langle E(Y|X_2), g_1(X_1) \rangle$$

$$+ \langle Y - E(Y|X_2), g_2(X_2) \rangle - \langle E(Y|X_1), g_2(X_2) \rangle$$

$$= 0 - 0 + 0 - 0$$

definition of P_1 Remark 1 below Definition of P_2 Remarks below.

Remark 1

$$\langle E(Y|X_2), g_1(X_1) \rangle = E(E(Y|X_2) \cdot g_1(X_1)) \underset{\text{of } X_1 \text{ and } X_2}{\overset{\text{Independence}}{=}}$$

$$= E(E(Y|X_2)) \cdot E(g_1(X_1))$$

$$= EY \cdot Eg_1(X_1) \underset{\text{Assumption}}{=} 0$$