

Supplemental Notes

EESD3 Week 09

Dr. Franzke

HW #9

① Leon-Garcia

5.75, 5.80, 6.50 ab

8.24 ab, 8.28 ab, 8.29 a

8.30 a

② Gubner

↑
(Practice, don't turn in)

7.35 - 7.39

("SIT" technique)

Topics

① Conditioning

- total expectation
- total variance
- total covariance

② "SIT" technique

③ Maximum Likelihood Estimation

Conditioning

Total Expectation

$$E_x[X] = E_y[E[X|Y]].$$

Total Variance

$$V_x[X] = E_y[V[X|Y]] + V_y[E[X|Y]].$$

Total Covariance

$$\begin{aligned} \text{Cov}(X, Y) &= E_z[\text{Cov}(X, Y|Z)] + \\ &\quad \text{Cov}_z(E[X|Z], E[Y|Z]) \end{aligned}$$

Total Probability : $P(B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)$

$$P(E) = \sum_k P(E|H_k) \cdot P(H_k) \quad \begin{array}{l} \text{if } \{H_k\} \\ \text{partition } \Omega \end{array}$$

Thm. $P(X \in A) = \int_{y=-\infty}^{y=\infty} P(X \in A | Y=y) \cdot f_Y(y) dy$
 $= E_Y[P(X \in A | Y)]$ c.v.

Prf. $P(X \in A) = P((X, Y) \in A \times \mathbb{R})$ if $A \subset \mathbb{R}$.

$$= \iint_{A \times \mathbb{R}} f_{X,Y}(x, y) dy dx$$

$$\begin{aligned}
 & \stackrel{\text{(Fubini)}}{=} \int_A \int_{y=-\infty}^{y=\infty} f_{XY}(x, y) \cdot dy dx \\
 & = \int_{y=-\infty}^{y=\infty} \left(\int_A f_{XY}(x, y) dx \right) f_Y(y) dy. \quad \text{since } f(x, y) = \frac{f(x, y)}{f(y)} \\
 & = \int_{y=-\infty}^{y=\infty} P[X \in A \mid Y=y] f_Y(y) dy.
 \end{aligned}$$

QED

Similarly:

$$\underline{\text{Thm:}} \quad P[(X, Y) \in B] = \int_{x=-\infty}^{x=\infty} P[(X, Y) \in B \mid X=x] \cdot f_X(x) dx.$$

★ "SIT" technique for solving problems

- (1) T: use Total probability (or expectation/variance/etc.)
- (2) S: Substitute x for r.v. X if $X=x$.
- (3) I: Use Independence of X and Y (if appropriate)

Ex: Independent X and Y : $X \sim U[0,1]$, $Y \sim \text{Exp}(1)$

find $P(Y > \ln(1+X))$

(use SET)

$$P(Y > \ln(1+X)) = P((X, Y) \in B)$$

$$\stackrel{\text{Total prob}}{=} \int_{x=-\infty}^{x=\infty} P(Y > \ln(1+X) | X=x) f_X(x) dx$$

$$\stackrel{X \sim U[0,1]}{=} \int_0^1 P(Y > \ln(1+X) | X=x) dx$$

$$\stackrel{\text{Sub}}{=} \int_0^1 P(Y > \ln(1+x) | X=x) dx$$

$$\stackrel{\text{Ind}}{=} \int_0^1 P(Y > \ln(1+x)) dx$$

$$\stackrel{Y \sim \text{Exp}(1)}{=} \int_0^1 e^{-\ln(1+x)} dx$$

$$= \int_0^1 e^{\ln(\frac{1}{1+x})} dx$$

$$= \int_0^1 \frac{dx}{1+x}$$

put $u=1+x$
 $du=dx$

$x=0 \rightarrow u=1$
 $x=1 \rightarrow u=2$

$$\therefore \stackrel{u=1+x}{=} \int_1^2 \frac{du}{u} = \ln u \Big|_1^2 = \ln 2 - \cancel{\ln 1} = \underline{\ln 2} \approx 0.693$$

Defn: Conditional Expectation r.v. $E[Y|X]$:

$$T_X: E[Y|X=x] = \begin{cases} \int_{y=-\infty}^{y=\infty} y \cdot f_{Y|X}(y|x) dy & \text{continuous} \\ \sum_y y \cdot P_{Y|X}(y|x) & \text{discrete} \end{cases}$$



Note: $E[Y|X=x] = g(x)$ $\uparrow$ constant $\therefore g(x) = E[Y|X]$ $\uparrow$ random variable

Fact: formally $E[Y|X]$ arises from Radon-Nikodym Theorem

$$\therefore E[I_B | I_A] = P(B|A).$$

Thm: (Total Expectation)

$$E_x[X] = E_y[E[X|Y]] \quad (\text{it exists})$$

$$\begin{aligned} \text{Prf: } E_y[E[X|Y]] & \stackrel{\text{cont.}}{=} \int_{y=-\infty}^{y=\infty} E[X|Y=y] \cdot f_y(y) dy \quad (\text{assume exists}) \\ & = \int_{y=-\infty}^{y=\infty} \int_{x=-\infty}^{x=\infty} x \cdot f(x|y) \cdot dx \cdot f_y(y) dy \\ & \stackrel{\text{Fubini}}{=} \int_{x=-\infty}^{x=\infty} x \left[\int_{y=-\infty}^{y=\infty} f_{xy}(x,y) dy \right] dx \quad \text{since } f(x|y) = \frac{f(x,y)}{f(y)} \\ & = \int_{x=-\infty}^{x=\infty} x \cdot f_x(x) dx \\ & = E_x[X] \end{aligned}$$

QED

Similarly:

$$(1) E_{xy}[g(x,y)] = \int_{x=-\infty}^{x=\infty} E[g(x,y) | X=x] \cdot f_x(x) dx$$

$$(2) E_{xyz}[g(x,y,z)] = \int_{z=-\infty}^{z=\infty} \int_{y=-\infty}^{y=\infty} E[g(x,y,z) | Y=y, Z=z] \cdot f_{yz}(y,z) dy dz$$

Ex: Dependent X and Y : $X \sim \text{Exp}(1)$, $Y_{|X=x} \sim N(0, x^2)$

$\uparrow$
 $x > 0$

Find $E[Y^2 X^4]$

use TS of "SIT"

$$\therefore E[Y^2 X^4] \stackrel{\text{total exp.}}{=} \int_{x=-\infty}^{x=\infty} E[Y^2 X^4 | X=x] f_x(x) dx.$$

$E_x[E[Y^2 X^4 | x]]$

$$\stackrel{X \sim \text{Exp}(1)}{=} \int_0^{\infty} E[Y^2 X^4 | X=x] e^{-x} dx$$

$$\stackrel{\text{Sub}}{=} \int_0^{\infty} E[Y^2 x^4 | X=x] e^{-x} dx$$

$$= \int_0^{\infty} x^4 \cdot E[Y^2 | X=x] e^{-x} dx \quad (\text{Proposition 1, below})$$

$$= \int_0^{\infty} x^4 x^2 e^{-x} dx \quad \text{since } Y_{|X=x} \sim N(0, x^2)$$

$$= \int_0^{\infty} x^6 e^{-x} dx$$

$$= E_x[X^6]$$

$$= \underline{6!} = 720$$

Since $X \sim \text{Exp}(1)$

General Case: S.A. $\mathcal{B} \subset \mathcal{A}$ S.A.

$$E[E[X | \mathcal{A}] | \mathcal{B}] = E[X | \mathcal{B}].$$

Proposition 1: $E[g(X)Y|X] = g(X) \cdot E[Y|X]$.

Prf: $\forall x$:

$$\begin{aligned} E[g(X)Y|X=x] &= \int_{y=-\infty}^{y=\infty} g(x) \cdot y \cdot f(y|x) dy \\ &= g(x) \cdot \int_{y=-\infty}^{y=\infty} y \cdot f(y|x) dy \\ &= g(x) \cdot E[Y|X=x]. \end{aligned}$$

$$\therefore E[g(X) \cdot Y|X] = g(X) \cdot E[Y|X].$$

Defn: Conditional Variance $V[Y|X] = E[(Y - E[Y|X])^2|X]$
(if exists)

Proposition 2: $V[Y|X] = E[Y^2|X] - E^2[Y|X]$

Prf: $V[Y|X] = E[(Y - E[Y|X])^2|X]$

$$\stackrel{\text{Lin}}{=} E[Y^2 + E^2[Y|X] - 2 \cdot Y E[Y|X] | X]$$

$$\stackrel{\text{Lin}}{=} E[Y^2|X] + E[\overbrace{E^2[Y|X]}^{g(x)} | X] - 2 \cdot E[Y \overbrace{E[Y|X]}^{h(x)} | X]$$

$$\stackrel{\text{prop 1}}{=} E[Y^2|X] + E^2[Y|X] \cdot E[1|X] - 2 \cdot E[Y|X] \cdot E[Y|X]$$

$$= E[Y^2|X] + E^2[Y|X] - 2 \cdot E^2[Y|X]$$

$$= E[Y^2|X] - E^2[Y|X]$$

QED

Fact: If X and Y are jointly Gaussian (JG)

$$f_{X,Y}(x,y) = \frac{\exp \left[- \frac{\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho_{xy} \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right)}{2(1-\rho_{xy}^2)} \right]}{2\pi \sigma_x \sigma_y \sqrt{1-\rho_{xy}^2}}$$

then:

memorize

$$(1) E[Y|X] = \mu_y + \rho_{xy} \frac{\sigma_y}{\sigma_x} (X - \mu_x)$$

$$(2) V[Y|X] = \sigma_y^2 (1 - \rho_{xy}^2)$$

$$\therefore \sigma_{Y|X}^2 < \sigma_y^2 \quad \text{if} \quad \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \neq 0.$$

Thm: (Total Variance)

$$V_X[X] = E_Y[V[X|Y]] + V_Y[E[X|Y]]. \quad \text{"ex-ve"}$$

Prf: $E_Y[V[X|Y]] + V_Y[E[X|Y]]$

$$= E_Y[V[X|Y]] + E_Y[E^2[X|Y]] - (E_Y[E[X|Y]])^2$$

total exp

$$= E_Y[V[X|Y]] + E_Y[E^2[X|Y]] - E_X^2[X].$$

prop 2

$$= E_Y[E[X^2|Y] - E^2[X|Y]] + E_Y[E^2[X|Y]] - E_X^2[X]$$

$$\begin{aligned}
&= E_Y[E(X^2|Y)] - \cancel{E_Y[E^2(X|Y)]} \\
&\quad + \cancel{E_Y[E^2(X|Y)]} - E_X^2[X] \\
&\stackrel{\text{total exp}}{=} E_X[X^2] - E_X^2[X] \\
&= V_X[X]
\end{aligned}$$

QED

★ Thm: (Doubly Random Sum) If iid $X_1, \dots, X_n$ $\forall n$
 and $\sigma_X^2 < \infty$ ($\because \mu_X < \infty$). — Positive $\mathbb{Z}^+$ r.v. N
 — N is independent of $X_1, \dots, X_n \forall n$.

$$(1) E\left[\sum_{k=1}^N X_k\right] = E_N[N] \cdot \mu_X$$

$$(2) V\left[\sum_{k=1}^N X_k\right] = E_N[N] \sigma_X^2 + V_N[N] \mu_X^2$$

Proof:

(SIT technique)

$$(1) E\left[\sum_{k=1}^N X_k\right] \stackrel{\text{total exp}}{=} E_N\left[E\left[\sum_{k=1}^N X_k | N\right]\right].$$

$$= \sum_{n=1}^{\infty} p_N(n) \cdot E\left[\sum_{k=1}^N X_k | N=n\right]. \quad \text{since } N > 0$$

$$\stackrel{\text{sub}}{=} \sum_{n=1}^{\infty} p_N(n) E\left[\sum_{k=1}^n X_k | N=n\right]$$

$$\stackrel{\text{ind}}{=} \sum_{n=1}^{\infty} p_N(n) \cdot E\left[\sum_{k=1}^n X_k\right]$$

$$\stackrel{\text{lin}}{=} \sum_{n=1}^{\infty} p_N(n) \cdot \sum_{k=1}^n E[X_k].$$

$$\begin{aligned}
& \stackrel{\text{iid}}{=} \sum_{n=1}^{\infty} p_N(n) (n \cdot \mu_x) \\
& = \mu_x \cdot \sum_{n=1}^{\infty} n \cdot p_N(n) \\
& = \mu_x E_N[N] = E_N[N] \cdot \mu_x
\end{aligned}$$

QED.

$$\begin{aligned}
(2) \quad V\left[\sum_{k=1}^N X_k\right] & \stackrel{\text{total var}}{=} E_N\left[V\left[\sum_{k=1}^N X_k \mid N\right]\right] + V_N\left[E\left[\sum_{k=1}^N X_k \mid N\right]\right] \\
& = \sum_{n=1}^{\infty} p_N(n) \cdot V\left[\sum_{k=1}^N X_k \mid N=n\right] + V_N\left[E\left[\sum_{k=1}^N X_k \mid N\right]\right] \\
& \stackrel{\text{sub}}{=} \sum_{n=1}^{\infty} p_N(n) \cdot V\left[\sum_{k=1}^n X_k \mid N=n\right] + V_N\left[E\left[\sum_{k=1}^N X_k \mid N\right]\right] \\
& \stackrel{\text{ind}}{=} \sum_{n=1}^{\infty} p_N(n) \cdot V\left[\sum_{k=1}^n X_k\right] + V_N\left[E\left[\sum_{k=1}^N X_k \mid N\right]\right] \\
& \stackrel{\text{iid}}{=} \sigma_x^2 \sum_{n=1}^{\infty} n \cdot p_N(n) + V_N\left[E\left[\sum_{k=1}^N X_k \mid N\right]\right] \\
& = E_N[N] \sigma_x^2 + V_N\left[E\left[\sum_{k=1}^N X_k \mid N\right]\right] \\
& = E_N[N] \sigma_x^2 + E_N\left[\left(E\left[\sum_{k=1}^N X_k \mid N=n\right]\right)^2\right] \\
& \quad - \left(E_N\left[E\left[\sum_{k=1}^N X_k \mid N=n\right]\right]\right)^2 \\
& \stackrel{\text{sub}}{=} E_N[N] \sigma_x^2 + E_N\left[\left(E\left[\sum_{k=1}^n X_k \mid N=n\right]\right)^2\right] \\
& \quad - \left(E_N\left[E\left[\sum_{k=1}^n X_k \mid N=n\right]\right]\right)^2 \\
& \stackrel{\text{ind}}{=} E_N[N] \sigma_x^2 + E_N\left[\left(E\left[\sum_{k=1}^n X_k\right]\right)^2\right] \\
& \quad - \left(E_N\left[E\left[\sum_{k=1}^n X_k\right]\right]\right)^2
\end{aligned}$$

$$\begin{aligned}
& \stackrel{\text{iid}}{=} E_N[N] \sigma_x^2 + \mu_x^2 \sum_{n=1}^{\infty} n^2 \cdot p_N(n) - \mu_x^2 \left(\sum_{n=1}^{\infty} n \cdot p_N(n) \right)^2 \\
& = E_N[N] \sigma_x^2 + \mu_x^2 \cdot E_N[N^2] - \mu_x^2 \cdot E_N^2[N] \\
& = E_N[N] \sigma_x^2 + \mu_x^2 (E_N[N^2] - E_N^2[N]) \\
& = E_N[N] \sigma_x^2 + \mu_x^2 V_N[N].
\end{aligned}$$

QED

Similarly:

★ Thm: ("Doubly Random Sample Mean") If

- iid $X_1, \dots, X_n$ for $\forall n$ and $\sigma_x^2 < \infty$ ($\therefore \mu_x < \infty$)
- positive $\mathbb{R}^+$ -valued r.v. N
- N is independent of $X_1, \dots, X_n$ for $\forall n$:

$$(1) E[\bar{X}_N] = \mu_x$$

$$(2) V[\bar{X}_N] = \sigma_x^2 \cdot E_N\left[\frac{1}{N}\right]$$

Prf: (at home)

Proposition 3: $\text{Cov}(X, Y | Z) = E[XY | Z] - E[X | Z] \cdot E[Y | Z]$

★ Thrm: ("Total Covariance")

$$\text{Cov}_{XY}[X, Y] = E_z[\text{Cov}(X, Y|Z)] + \text{Cov}_{ZZ}[E[X|Z], E[Y|Z]]$$

Prf:

$$E_z[\text{Cov}(X, Y|Z)] + \text{Cov}_{ZZ}[E[X|Z], E[Y|Z]].$$

$$\stackrel{\text{prop 3}}{=} E_z[E[XY|Z]] - E_z[E[X|Z] \cdot E[Y|Z]] + \text{Cov}_{ZZ}[E[X|Z], E[Y|Z]]$$

$$= \underbrace{E_{XY}[XY]}_{E_z[E[XY|Z]]} - E_z[E[X|Z] \cdot E[Y|Z]] + E_z[(E[X|Z] - \underbrace{E_z[E[X|Z]]}_{E[X]})(E[Y|Z] - \underbrace{E_z[E[Y|Z]]}_{E[Y]})]$$

$$\stackrel{\text{total exp}}{=} E_{XY}[XY] - E_z[E[X|Z] \cdot E[Y|Z]] + E_z[(E[X|Z] - E_X[X])(E[Y|Z] - E_Y[Y])]$$

$$\stackrel{\text{lin}}{=} E_{XY}[XY] - E_z[E[X|Z] \cdot E[Y|Z]] + E_z[E[X|Z] \cdot E[Y|Z] - E_Y[Y] \cdot E[X|Z] - E_X[X] E[Y|Z] + E_X[X] \cdot E_Y[Y]]$$

$$\stackrel{\text{lin}}{=} E_{XY}[XY] - \cancel{E_z[E[X|Z] \cdot E[Y|Z]]} + \underbrace{E_X[X]}_{E_Y[Y]} \cdot \underbrace{E_z[E[X|Z]]}_{E_Y[Y]} - E_X[X] \cdot \underbrace{E_z[E[Y|Z]]}_{E_Y[Y]} + E_X[X] \cdot E_Y[Y].$$

$$\begin{aligned}
 & \stackrel{\text{total exp}}{=} E_{XY}[XY] - E_Y[Y] \cdot E_X[X] - \cancel{E_Y[Y]E_X[X]} + \cancel{E_Y[Y] \cdot E_X[X]} \\
 & = E_{XY}[XY] - E_X[X] \cdot E_Y[Y] \\
 & = \text{Cov}_{XY}(X, Y)
 \end{aligned}$$

QED

Recall: Central Limit Theorem (CLT) if iid $X_1, \dots, X_n$ (i.i.d.)

and $\sigma_x^2 < \infty$:

$$Z_n \triangleq \text{STD}(\bar{X}_n) = \frac{\bar{X}_n - \mu_x}{\sigma_x / \sqrt{n}} = \frac{\sum_{k=1}^n X_k - n \cdot \mu_x}{\sqrt{n} \cdot \sigma_x} \xrightarrow{d} Z \sim N(0, 1)$$

$\therefore$ For "large" n ($n \geq 30$):

$$Z_n \stackrel{d}{\approx} Z \sim N(0, 1)$$

$$\longleftrightarrow \frac{\bar{X}_n - \mu_x}{\sigma_x / \sqrt{n}} \stackrel{d}{\approx} Z$$

$$\longleftrightarrow \bar{X}_n \stackrel{d}{\approx} \mu_x + \frac{\sigma_x}{\sqrt{n}} Z \sim N\left(\mu_x, \frac{\sigma_x^2}{n}\right)$$

Because $E[\bar{X}_n] = \mu_x$ and $V[\bar{X}_n] = \frac{\sigma_x^2}{n}$

(and later) $a + bZ \sim N$ if $Z \sim N$

$\therefore$ For large n ($n \geq 30$):

$$\star \quad \boxed{\bar{X}_n \approx N\left(\mu_x, \frac{\sigma_x^2}{n}\right)} \longrightarrow \delta\text{-spike at } \mu_x$$

Similarly: $\sum_{k=1}^n X_k \approx N(n \cdot \mu_x, n \cdot \sigma_x^2) \longrightarrow \text{"improper" } N(\pm\infty, \pm\infty)$

Q: what if $n < 30$? ("small" n)

Thrm: iid $X_1, \dots, X_n \sim N(\mu_x, \sigma_x^2)$:

$$(1) \bar{X}_n \sim N\left(\mu_x, \frac{\sigma_x^2}{n}\right)$$

$$(2) \frac{(n-1)S_x^2}{\sigma_x^2} \sim \chi^2(n-1) \quad \text{for } S_x^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

★ (3) $\bar{X}_n$ and S_n^2 are independent
(not true in general)

Defn: t-distribution $T \sim t(r)$ and $r \geq 1$

$$f_x(x) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi r} \Gamma\left(\frac{r}{2}\right) \left(1 + \frac{x^2}{r}\right)^{\frac{r+1}{2}}} \quad \swarrow \text{bell curve}$$

Facts: (1) $r=1$: $T \sim C(0,1)$ standard Cauchy

$$(2) r > 1: E_T[T] = 0$$

$$(3) r > 2: V_T[T] = \frac{r}{r-2}$$

$$(4) r \uparrow \infty: \underline{T \xrightarrow{\text{d}} Z} \sim N(0,1)$$

Thrm: $T = \frac{Z}{\sqrt{U/r}} \sim t(r)$ if $Z \sim N(0,1)$ independent $U \sim \chi^2(r)$

Thrm: (Student's Theorem) iid $X_1, \dots, X_n \sim N(\mu_x, \sigma_x^2)$

$$T \stackrel{\Delta}{=} \frac{\bar{X}_n - \mu_x}{S_n / \sqrt{n}} \sim t(n-1)$$

Prf: $- Z_n = \frac{\bar{X}_n - \mu_x}{S_x / \sqrt{n}} \sim N(0,1)$ since $\bar{X}_n \sim N(\mu_x, \frac{\sigma_x^2}{n})$
 $\uparrow$ iid $X_k \sim N(\mu_x, \sigma_x^2)$

$- \frac{(n-1)}{S_x^2} \cdot S_n^2 \sim \chi^2(n-1)$ since iid $X_k \sim N(\mu_x, \sigma_x^2)$

$- Z_n$ and S_n^2 independent since $\bar{X}_n$ and S_n^2 independent.

$$\begin{aligned} \therefore T &= \frac{Z_n}{\sqrt{\left(\frac{(n-1)}{S_x^2} S_n^2\right) / (n-1)}} \sim t(n-1) \\ &= \frac{\left(\frac{\bar{X}_n - \mu_x}{S_x / \sqrt{n}}\right)}{\sqrt{\left(\frac{(n-1)}{S_x^2} S_n^2\right) / (n-1)}} \\ &= \frac{\bar{X}_n - \mu_x}{S_n / \sqrt{n}} \end{aligned}$$

QED.

$\therefore$ Rule of Thumb use $T(n-1)$ instead of Z if:

①

(1) $n < 30$

and

(at least approximately)

(2) data $x_1, \dots, x_n$ unimodal/symmetric

$$T = \frac{\bar{X}_n - \mu_x}{S_n / \sqrt{n}} \quad \text{vs.} \quad Z = \frac{\bar{X}_n - \mu_x}{\sigma_x / \sqrt{n}}$$

OR

②

if don't know σ_x .

Else: use non-parametric techniques (such as BOOTSTAPS)

Covariance Matrices

$$K_{xx} = E_x \left[(X - E_x[X]) (X - E_x[X])^T \right] = K_{xx}^T$$

$n \times n$ $n \times 1$

- real and symmetric ($\therefore$ e.v. $\lambda_k \in \mathbb{R}$)

$\therefore$ Diagonalizable $K_{xx} = O \Lambda O^T$ for $O O^T = I$.

$$\forall w \neq 0: w^T K_{xx} w = V_x [w^T X] \geq 0$$

$\therefore K_{xx}$ is positive semi-definite (PSD) $\therefore$ eigenvalues $\lambda_k \geq 0$

Vector / Matrix LLN

$$\bar{X}_n \xrightarrow{\text{op}} \mu_x$$

$n \times 1$

$$\hat{K}_n = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)(X_k - \bar{X}_n)^T \xrightarrow{\bullet} K_{xx}$$

(if finite $E[X_k^2]$)

Fact: If $Y = AX$

$n \times 1$ $n \times n$ $n \times 1$

$$(1) E_Y[Y] = A E_x[X]$$

$$(2) K_{YY} = A \cdot K_{xx} A^T$$

Q: How to "whiten" a K_{xx} ?

Given $X \sim K$ Find A such that if $Y = AX$

$n \times 1$ $n \times n$

then $K_{YY} = I$

$\therefore K_{YY}$ is "white"

$\uparrow$ uncorrelated $Y_1, \dots, Y_n$

Thm: If $X \sim K_{xx}$ and $Y = AX$ then $A = \Lambda^{-\frac{1}{2}} O^T$

"whitens" K_{xx} : $K_{yy} = I$ where $K_{xx} e_k = \lambda_k e_k$ and $\lambda_k > 0$

$\forall k$. $\Lambda = \text{Diagonal}(\lambda_1, \dots, \lambda_n)$

$O = E$ normalized for $E = [e_1 | e_2 | \dots | e_n]$
 $\therefore O O^T = I$.

Prf: K_{xx} positive definite (since $\lambda_k > 0$)

$$\therefore K_{xx} = O \Lambda O^T$$

$$\therefore K_{yy} = A K_{xx} A^T$$

$$= \Lambda^{-\frac{1}{2}} \underbrace{O^T O}_{I} \underbrace{O O^T}_{I} \Lambda^{-\frac{1}{2}}$$

since $O = O^T$ if O diagonal

$$= \Lambda^{-\frac{1}{2}} \Lambda \Lambda^{-\frac{1}{2}}$$

$$= \Lambda (\Lambda^{-\frac{1}{2}} \Lambda^{-\frac{1}{2}})$$

since diagonal matrices commute

$$= \Lambda \Lambda^{-1}$$

$$= I$$

QED

Special Case: (2x2 matrices)

$$K_{xx} = \begin{matrix} x & y \\ y & x \end{matrix} \begin{pmatrix} \overset{x}{\sigma_x^2} & \overset{y}{\sigma_{xy}} \\ \sigma_{xy} & \overset{y}{\sigma_y^2} \end{pmatrix} = \begin{matrix} x & y \\ y & x \end{matrix} \begin{pmatrix} \overset{x}{\sigma_x^2} & \overset{y}{\rho_{xy} \sigma_x \sigma_y} \\ \rho_{xy} \sigma_x \sigma_y & \overset{y}{\sigma_y^2} \end{pmatrix}$$

$$\forall A \in \mathbb{C}^{2 \times 2} \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$$\text{Tr}(A) = \text{Trace}(A) = \sum_i a_{ii} = a + d$$

$$D(A) = \text{Determinant}(A) = ad - bc$$

$$A^{-1} \text{ exists iff } D(A) \neq 0: A^{-1} = \frac{1}{D} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$$\therefore \text{Det}(K_{xx}) = \sigma_x^2 \sigma_y^2 - \rho_{xy}^2 \sigma_x^2 \sigma_y^2 = \sigma_x^2 \sigma_y^2 (1 - \rho_{xy}^2)$$

$$\therefore K_{xx}^{-1} = \frac{1}{\sigma_x^2 \sigma_y^2 (1 - \rho_{xy}^2)} \begin{pmatrix} \sigma_y^2 & -\rho_{xy} \sigma_x \sigma_y \\ -\rho_{xy} \sigma_x \sigma_y & \sigma_x^2 \end{pmatrix}$$

$$\text{if } -1 < \rho_{xy} < 1 \quad \left(\overset{\text{u.p.}}{\sigma_{xy}^2 \leq \sigma_x^2 \sigma_y^2} \right)$$

Fact: (1) $D(A) = \prod_{k=1}^n \lambda_k$

(2) $T(A) = \sum_{k=1}^n \lambda_k$

for A 's eigenvalues $\lambda_1, \dots, \lambda_n$: $A e_k = \lambda_k e_k \quad (e_k \neq 0)$.

Characteristic equation in 2×2 case

$$\lambda^2 - T\lambda + D = 0$$

$$\therefore \lambda = \frac{T \pm \sqrt{T^2 - 4D}}{2} \quad \text{memorize} \quad \star$$

$$\therefore A - \lambda I = \begin{pmatrix} u & v \\ cu & cv \end{pmatrix} \quad \text{some constant } c$$

$$\therefore \text{eigenvector } e_\lambda = \begin{pmatrix} -v \\ u \end{pmatrix}$$

Ex: Whiten $K_{xx} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$

$$T=4$$

$$D=3$$

$$\therefore \lambda = \frac{T \pm \sqrt{T^2 - 4D}}{2}$$

$$= \frac{4 \pm \sqrt{16 - 12}}{2} = \frac{4 \pm 2}{2} = 1 \text{ or } 3$$

Say $\lambda_1 = 1$ and $\lambda_2 = 3$

$$\therefore \lambda_1 = 1: K_{xx} - I = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ (-1) \cdot 1 & (-1) \cdot (-1) \end{pmatrix}$$

$$\therefore e_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \therefore \text{normalize } u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$\uparrow$
 $\frac{e}{\|e\|}$

$$\lambda_2 = 3: K_{xx} - 3I = \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix}$$

$$\therefore e_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \therefore u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$\therefore \Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \quad \text{and} \quad \Lambda^{-\frac{1}{2}} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sqrt{3}} \end{pmatrix}$$

$$O = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = O^T \quad (O^T \cdot O = I)$$

$$\therefore O \Lambda O^T = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

$\nearrow$
 $= K_{xx}!$

$$\therefore \text{Put } A = \Lambda^{-\frac{1}{2}} O^T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{pmatrix}$$

$$\begin{aligned}
\therefore K_{YY} &= A K_{XX} A^T \\
&= \frac{1}{2} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ 1 & -\frac{1}{\sqrt{2}} \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.
\end{aligned}$$

$\therefore K_{YY}$ is "white"

$\therefore Y_1$ and Y_2 are uncorrelated

where $\begin{cases} Y_1 = \frac{1}{\sqrt{2}} (X_1 + X_2) \\ Y_2 = \frac{1}{\sqrt{2}} (X_1 - X_2) \end{cases}$

from $Y = (A^{-\frac{1}{2}} \Sigma^T) X$

Optimal Estimators

Minimum Mean-Squared $\hat{\Theta}_n^{ms}$

Fact: $\hat{\Theta}_n^{ms} = E[Y|X]$

Minimum Absolute Error $\hat{\Theta}_n^{abs}$

$$\hat{\Theta}_n^{abs} = \text{Median}$$

Bayesian estimator

- MAP: Maximum A Posteriori $\hat{\Theta}_n^{MAP}$

$$\hat{\Theta}_n^{MAP} = \arg \max_{\Theta} f(\Theta|x)$$

$$\text{where } f(\Theta|x) \propto \underbrace{h(\Theta)}_{\text{prior}} \cdot \underbrace{g(x|\Theta)}_{\text{likelihood}}$$

$$\text{and } \Theta \sim h(\Theta)$$

Maximum Likelihood

$$\hat{\Theta}_n^{ML} = \arg \max_{\Theta} g(x|\Theta) \quad \Theta \in \Theta$$

- weakness: requires functional form g
(else use mode of empirical pdf)
- leads to iterative EM (expectation-maximization algorithm)
- MLE = MAP if $h(\Theta) = \text{Uniform.}$

MLE properties

① Consistency $\hat{\theta}_n^{ML} \xrightarrow{P} \theta$ (need smoothness conditions)

② Asymptotic Normality (from CLT and LLN)

$$\sqrt{n} (\hat{\theta}_n^{ML} - \theta) \xrightarrow{d} W \sim N(0, \frac{1}{I(\theta)})$$

I = Fisher Information

③ Invariance Principle (I.P.) ★

$$\widehat{g(\theta)}^{ML} \stackrel{\text{I.P.}}{=} g(\hat{\theta}_n^{ML}) \text{ for any function } g.$$

(false for $\hat{\theta}^{MS}$ in general)

Log-likelihood Function L :

$$L \triangleq \ln f(x_1, \dots, x_n | \theta) \quad \text{r.s. } X_1, \dots, X_n \sim \theta$$

$$\stackrel{\text{iid}}{=} \ln \prod_{k=1}^n f(x_k)$$

$$\stackrel{\text{log}}{=} \sum_{k=1}^n \ln f(x_k)$$

Now find MODE of L by any means, e.g.

- combinatorial optimization

- differentiation (if possible)

Differential Case: ("null the gradient")

(1) Set $\frac{dL}{d\theta} = 0$

(2) Solve (1) for $\hat{\theta}_n^{ML}$

(3) Check maximization $\frac{d^2 L}{d\theta^2} < 0$ at $\hat{\theta}_n^{ML}$ value

Ex: Find $\hat{\theta}_n^{ML}$ if iid $X_1, \dots, X_n \sim \exp(\theta)$

$$\therefore f_{\theta}(x) = f(x|\theta) = \frac{1}{\theta} e^{-x/\theta} \text{ for } x > 0 \text{ and } \theta > 0$$

$$L = \ln f_{\theta}(x_1, \dots, x_n) \stackrel{\text{iid}}{=} \sum_{k=1}^n \ln f_{\theta}(x_k)$$

$$= \sum_{k=1}^n \ln \frac{e^{-x_k/\theta}}{\theta} = \sum_{k=1}^n \left(-\frac{x_k}{\theta} - \ln \theta \right)$$

$$= -\frac{1}{\theta} \sum_{k=1}^n x_k - n \cdot \ln \theta$$

$$\therefore \frac{dL}{d\theta} = 0 \text{ gives:}$$

$$0 = \frac{1}{(\hat{\theta}_n^{ML})^2} \sum_{k=1}^n x_k - \frac{n}{\hat{\theta}_n^{ML}} \quad \left(\frac{d^2 L}{d\theta^2} < 0 \right).$$

$$\longleftrightarrow n \cdot \hat{\theta}_n^{ML} = \sum_{k=1}^n x_k$$

$$\longleftrightarrow \hat{\theta}_n^{ML} = \frac{1}{n} \cdot \sum_{k=1}^n x_k$$

$$\therefore \hat{\theta}_n^{ML} = \overline{X}_n$$

Ex: Normal case.

$$\text{Find } \hat{\Theta}_n^{ML} = \begin{cases} \hat{\mu}_n^{ML} \\ (\hat{\Sigma}_n^2)^{ML} \end{cases} \quad \text{if iid } X_1, \dots, X_n \sim N(\mu_x, \Sigma_x^2)$$

Set $\nabla_{\Theta} L = 0$ and solve to give

$$\hat{\Theta}_n^{ML} = \begin{cases} \hat{\mu}_n^{ML} = \bar{X}_n \\ (\hat{\Sigma}_n^2)^{ML} = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \\ = \frac{n-1}{n} S_n^2 \end{cases} \quad \begin{array}{l} \swarrow \text{biased} \\ \text{(asymptotically unbiased)} \end{array}$$

Thm: (MLE Invariance Principle, I.P.)

$$\widehat{g(\Theta)}^{ML} \stackrel{\text{I.P.}}{=} g(\hat{\Theta}^{ML})$$

Prf: (only for case when g is 1-1, $u \stackrel{\text{I.P.}}{=} g(\Theta)$)

$$\begin{aligned} \therefore \max_{\Theta} L(g(\Theta)) &= \max_{u=g(\Theta)} L(u) \\ &\stackrel{\text{I.P.}}{=} \max_u L(g^{-1}(u)) \end{aligned}$$

maximum occurs when $g^{-1}(u) = \hat{\Theta}^{ML}$

$$\therefore \text{Take } \hat{u}^{ML} \stackrel{\text{I.P.}}{=} g(\hat{\Theta}^{ML})$$

QED

Ex: find $\hat{e}^{\cos \Theta}_{ML}$ if iid $X_k \sim \exp(\Theta)$

$$\begin{aligned}\therefore \hat{e}^{\cos \Theta}_{ML} &\stackrel{\text{I.P.}}{=} e^{\cos \hat{\Theta}_n^{ML}} \\ &= e^{\cos \bar{X}_n}\end{aligned}$$

since iid $X_k \sim \exp(\Theta)$

$$\therefore \hat{\Theta}_n^{ML} = \bar{X}_n$$