Supplemental Notes

EE503 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 7.35 – 7.39
   ("SIT" technique)
   (Practice, don't turn in)

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

$$\mathrm{Cov}(X,Y) = E_Z[\mathrm{Cov}(X,Y|Z)] + \mathrm{Cov}_Z(E[X|Z], E[Y|Z])$$

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

$$P(E) = \sum_k P(E|H_k)\cdot P(H_k)$$ if $\{H_k\}$ partition Ω

Thrm: $$P(X \in A) = \int_{y=-\infty}^{y=\infty} P(X\in A \mid Y=y)\cdot f_Y(y)\, dy$$

r.v.

$$= E_Y[P(X\in A \mid Y)]$$

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

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

$$\overset{\text{(Fubini)}}{=} \int_A \int_{y=-\infty}^{y=\infty} f_{XY}(x,y)\cdot dy\,dx$$

since $f(x|y) = \frac{f(x,y)}{f(y)}$

$$= \int_{y=-\infty}^{y=\infty} \left(\int_A f_{X|Y}(x|y)\,dx \right) f_Y(y)\,dy.$$

$$= \int_{y=-\infty}^{y=\infty} P[X \in A \mid Y = y]\, f_Y(y)\,dy.$$

QED

Similarly:

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

★ "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 SIT)

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

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

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

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

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

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

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

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

put $u = 1+x$, $du = dx$ $\therefore$ $x=0 \longrightarrow u=1$, $x=1 \longrightarrow u=2$

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

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

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

☆☆ Note: $E[Y \mid X=x] = g(x)$ ← constant $\therefore g(X) = E[Y \mid X]$. ← random variable

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

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

Thrm: (Total Expectation)

$$E_x[X] = E_Y[E[X|Y]] \qquad \text{(if exists)}$$

Prf: $E_Y[E[X|Y]] \overset{\text{cont.}}{=} \int_{y=-\infty}^{y=\infty} E[Y|X=x]\cdot f_Y(y)\,dy$ (assume exists)

$$= \int_{y=-\infty}^{y=\infty}\int_{x=-\infty}^{x=\infty} x\cdot f(x|y)\cdot f_Y(y)\,dy$$

$$\overset{\text{Fubini}}{=} \int_{x=-\infty}^{x=\infty} x\left[\int_{y=-\infty}^{y=\infty} f_{XY}(x,y)\,dy\right]dx \qquad \text{since } f(x|y)=\frac{f(x,y)}{f(y)}$$

$$= \int_{x=-\infty}^{x=\infty} x\cdot f_X(x)\,dx$$

$$= E_x[X]$$

QED

Similarly:

(1) $E_{XY}[g(X,Y)] = \int_{x=-\infty}^{x=\infty} E[g(X,Y) \mid X=x]\cdot f_x(x)\,dx$

(2) $E_{XYZ}[g(X,Y,Z)] =$

$$\int_{z=-\infty}^{z=\infty}\int_{y=-\infty}^{y=\infty} E[g(X,Y,Z) \mid Y=y, Z=z]\cdot f_{YZ}(y,z)\,dy\,dz$$

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

$\uparrow$ $x>0$

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

use TS of "SIT"

$\therefore$ $E[Y^2X^4] \overset{\text{total exp.}}{=} \int_{x=-\infty}^{x=\infty} E[Y^2X^4 \mid X=x] f_X(x)\, dx.$

$E_X[E[Y^2X^4 \mid X]]$

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

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

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

$= \int_0^\infty x^4 x^2 e^{-x}\, dx$ 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 $X \sim Exp(1)$

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

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

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

Prf: $\forall x$:

$$E[g(X)Y \mid X=x] = \int_{y=-\infty}^{y=\infty} g(x) \cdot y \cdot f(y \mid x)\, dy$$

$$= g(x) \cdot \int_{y=-\infty}^{y=\infty} y \cdot f(y \mid x)\, dy$$

$$= g(x) \cdot E[Y \mid X=x].$$

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

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

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

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

$$\overset{\text{lin}}{=} E\left[Y^2 + E^2[Y \mid X] - 2 \cdot Y\, E[Y \mid X] \,\middle|\, X \right].$$

$$\overset{\text{lin}}{=} E[Y^2 \mid X] + E[\overbrace{E^2[Y \mid X]}^{g(X)} \mid X]$$

$$- 2 \cdot E[Y \overbrace{E[Y \mid X]}^{h(X)} \mid X]$$

$$\overset{\text{prop 1}}{=} E[Y^2 \mid X] + E^2[Y \mid X] \cdot E[1 \mid X]$$

$$- 2 \cdot E[Y \mid X] \cdot E[Y \mid X]$$

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

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

QED

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

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

then:

memorize

(1) $E[Y|X] = \mu_Y + \rho_{xY}\frac{\sigma_Y}{\sigma_x}(X-\mu_x)$

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

$\therefore \sigma^2_{Y|x} < \sigma_Y^2$ if $\rho_{xY} = \frac{\sigma_{xY}}{\sigma_x\sigma_Y} \neq 0$.

Thrm: (Total Variance)

$$V_X[X] = E_Y[V[X|Y]] + V_Y[E[X|Y]].$$

"ev-ve"

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

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

$$\overset{\text{total exp}}{=} E_Y[V[X|Y]] + E_Y[E^2(X|Y)] - E_X^2[X].$$

$$\overset{\text{prop 2}}{=} E_Y[E[X^2|Y] - E^2[X|Y]] + E_Y[E^2(X|Y)] - E_X^2[X]$$

$$= E_Y[E[X^2|Y]] - \cancel{E_Y[E^2[X|Y]]}$$

$$+ \cancel{E_Y[E^2[X|Y]]} - E_X^2[X]$$

$$\overset{\text{total exp}}{=} E_X[X^2] - E_X^2[X]$$

$$= V_X[X]$$

QED

Thm: (Doubly Random Sum) If iid $X_1, \dots, X_n \ \forall n$ and $\sigma_X^2 < \infty$ ($\therefore \mu_X < \infty$).

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

Pf:

(SIT technique)

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

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

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

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

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

$$\overset{iid}{=} \sum_{n=1}^{\infty} p_N(n)\,(n \cdot \mu_x)$$

$$= \mu_x \cdot \sum_{n=1}^{\infty} n \cdot p_N(n)$$

$$= \mu_x E_N[N] = E_N[N] \cdot \mu_x$$

QED.

(2) $$V\left[\sum_{k=1}^{N} X_k\right] \overset{\text{total var}}{=} E_N\left[V\left[\sum_{k=1}^{N} X_k \,\middle|\, N\right]\right] + V_N\left[E\left[\sum_{k=1}^{N} X_k \,\middle|\, N\right]\right]$$

$$= \sum_{n=1}^{\infty} p_N(n) \cdot V\left[\sum_{k=1}^{N} X_k \,\middle|\, N=n\right] + V_N\left[E\left[\sum_{k=1}^{N} X_k \,\middle|\, N\right]\right]$$

$$\overset{sub}{=} \sum_{n=1}^{\infty} p_N(n) \cdot V\left[\sum_{k=1}^{n} X_k \,\middle|\, N=n\right] + V_N\left[E\left[\sum_{k=1}^{N} X_k \,\middle|\, N\right]\right]$$

$$\overset{ind}{=} \sum_{n=1}^{\infty} p_N(n) \cdot V\left[\sum_{k=1}^{n} X_k\right] + V_N\left[E\left[\sum_{k=1}^{N} X_k \,\middle|\, N\right]\right]$$

$$\overset{iid}{=} \sigma_x^2 \sum_{n=1}^{\infty} n \cdot p_N(n) + V_N\left[E\left[\sum_{k=1}^{N} X_k \,\middle|\, N\right]\right]$$

$$= E_N[N]\,\sigma_x^2 + V_N\left[E\left[\sum_{k=1}^{N} X_k \,\middle|\, N\right]\right]$$

$$= E_N[N]\,\sigma_x^2 + E_N\left[\left(E\left[\sum_{k=1}^{N} X_k \,\middle|\, N=n\right]\right)^2\right] - \left(E_N\left[E\left[\sum_{k=1}^{N} X_k \,\middle|\, N=n\right]\right]\right)^2$$

$$\overset{sub}{=} E_N[N]\,\sigma_x^2 + E_N\left[\left(E\left[\sum_{k=1}^{n} X_k \,\middle|\, N=n\right]\right)^2\right] - \left(E_N\left[E\left[\sum_{k=1}^{n} X_k \,\middle|\, N=n\right]\right]\right)^2$$

$$\overset{ind}{=} E_N[N]\,\sigma_x^2 + E_N\left[\left(E\left[\sum_{k=1}^{n} X_k\right]\right)^2\right] - \left(E_N\left[E\left[\sum_{k=1}^{n} X_k\right]\right]\right)^2$$

$$\overset{iid}{=} E_N[N]\sigma_X^2 + \mu_X^2 \sum_{n=1}^{\infty} n^2 \cdot p_N(n) - \mu_X^2 \left(\sum_{n=1}^{\infty} n \cdot p_N(n) \right)^2$$

$$= E_N[N]\sigma_X^2 + \mu_X^2 \cdot E_N[N^2] - \mu_X^2 \cdot E_N^2[N]$$

$$= E_N[N]\sigma_X^2 + \mu_X^2 \left(E_N[N^2] - E^2[N] \right)$$

$$= E_N[N]\sigma_X^2 + \mu_X^2 V_N[N].$$

QED

Similarly:

Thrm: ("Doubly Random Sample Mean") If

- iid $X_1, \dots, X_n$ for $\forall n$ and $\sigma_X^2 < \infty$ $(\therefore \mu_X < \infty)$
- positive $\mathbb{Z}^+$-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: $Cov(X, Y \mid Z) = E[XY \mid Z] - E[X \mid Z] \cdot E[Y \mid Z]$

Thrm: ("Total Covariance")

$$\mathrm{Cov}_{XY}[X,Y] = E_Z[\mathrm{Cov}[X,Y|Z]] + \mathrm{Cov}_{ZZ}[E[X|Z], E[Y|Z]]$$

Pf:

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

$$\overset{\text{prop3}}{=} E_Z[E[XY|Z]] - E[X|Z]\cdot E[Y|Z]$$

$$+ \mathrm{Cov}_{ZZ}[E[X|Z], E[Y|Z]]$$

$$= \overbrace{E_Z[E[XY|Z]]}^{E_{XY}[XY]} - E_Z[E[X|Z]\cdot E[Y|Z]]$$

$$+ E_Z[(E[X|Z] - \underbrace{E_Z[E[X|Z]]}_{E[X]})(E[Y|Z] - \underbrace{E_Z[E[Y|Z]]}_{E[Y]})].$$

$$\overset{\text{total exp}}{=} E_{XY}[XY] - E_Z[E[X|Z]\cdot E[Y|Z]]$$

$$+ E_Z[(E[X|Z] - E_X[X])(E[Y|Z] - E_Y[Y])]$$

$$\overset{\text{lin}}{=} E_{XY}[XY] - E_Z[E[X|Z]\cdot E[Y|Z]]$$

$$+ E_Z[E[X|Z]\cdot E[Y|Z] - E_Y[Y]\cdot E[X|Z]$$

$$- E_X[X]E[Y|Z] + E_X[X]\cdot E_Y[Y]]$$

$$\overset{\text{lin}}{=} E_{XY}[XY] - \cancel{E_Z[E[X|Z]\cdot E[Y|Z]]}.$$

$$+ \cancel{E_Z[E[X|Z]\cdot E[Y|Z]]} - E_Y[Y]\cdot \overbrace{E_Z[E[X|Z]]}^{E_X[X]}.$$

$$- E_X[X]\cdot \overbrace{E_Z[E[Y|Z]]}^{E_Y[Y]} + E_X[X]\cdot E_Y[Y].$$

$$\overset{\text{total exp}}{=} E_{XY}[XY] - E_Y[Y]\cdot E_X[X] - \cancel{E_Y[Y]E_X[X]} + \cancel{E_Y[Y]\cdot E_X[X]}$$

$$= E_{XY}[XY] - E_X[X]\cdot E_Y[Y]$$

$$= \mathrm{Cov}_{XY}(X,Y)$$

QED

<u>Recall</u>: Central Limit Theorem (CLT) if iid $X_1, \dots, X_n$ (r.s.)

and $\sigma_X^2 < \infty$:

$$Z_n \overset{\Delta}{=} \mathrm{STD}(\bar{X}_n) = \frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}} = \frac{\sum_{k=1}^{n} X_k - n\cdot\mu_X}{\sqrt{n}\cdot\sigma_X} \xrightarrow{d} Z \sim N(0,1)$$

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

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

$$\longleftrightarrow \frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}} \overset{d}{\approx} Z$$

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

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 ($n \geq 30$):

$$\boxed{X_n \approx N\left(\mu_X, \frac{\sigma_X^2}{n}\right)} \longrightarrow \delta\text{-spike at } \mu_X$$

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

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

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

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

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

★ (3) $\overline{X}_n$ and S_n^2 are independent

(not true in general)

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

$$f_x(x) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi r}\,\Gamma\left(\frac{r}{2}\right)\left(1+\frac{x^2}{r}\right)^{\frac{r+1}{2}}}$$

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

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

$$T \triangleq \frac{X_n - \mu_x}{S_n/\sqrt{n}} \sim t(n-1)$$

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

$-\ \frac{(n-1)}{\sigma_x^2} \cdot S_n^2 \sim \chi^2(n-1)$ 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.

$$\therefore\ T = \frac{Z_n}{\sqrt{\left(\frac{(n-1)}{\sigma_x^2} S_n^2\right) \Big/ (n-1)}} \sim t(n-1)$$

$$= \frac{\left(\frac{\bar{X}_n - \mu_x}{\cancel{\sigma_x}/\sqrt{n}}\right)}{\sqrt{\left(\frac{\cancel{(n-1)}}{\cancel{\sigma_x^2}} S_n^2\right) \Big/ \cancel{(n-1)}}}$$

$$= \frac{\bar{X}_n - \mu_x}{S_n/\sqrt{n}}$$

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

$$\underset{n\times n}{K_{XX}} = E_X\left[(\underset{n\times 1}{X - E_X[X]})(X - E_X[X])^T \right] = K_{XX}^T$$

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

$\therefore$ Diagonable $K_{XX} = O\Lambda O^T$ for $OO^T = I$.

$$\forall w \neq 0: \quad w^T K_{XX} w = V_X[w^T X] \geq 0$$

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

Vector | Matrix LLN

$$\underset{n\times 1}{\bar{X}_n} \xrightarrow{P} \mu_X$$

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

(if finite $E[X_k^4]$).

Fact: If $\underset{n\times 1}{Y} = \underset{n\times n}{A}\,\underset{n\times 1}{X}$

(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 $\underset{n\times 1}{X} \sim K$ **Find** $\underset{n\times n}{A}$ such that if $Y = AX$

then $K_{YY} = I$ $\therefore$ K_{YY} is "white"

uncorrelated $Y_1, \ldots, Y_n$

Thrm: If $\underset{n\times 1}{X} \sim \underset{n\times n}{K_{xx}}$ and $\underset{n\times 1}{Y} = \underset{n\times n}{A}\underset{n\times 1}{X}$ then $\boxed{A = \Lambda^{-\frac{1}{2}}\Theta^T}$

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

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

$\Theta = \underset{n\times n}{E}$ normalized for $E = [e_1 | e_2 | \cdots | e_n]$

$\therefore \Theta\Theta^T = I.$

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

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

$$\begin{aligned}
\therefore K_{YY} &= A K_{xx} A^T \\
&= \Lambda^{-\frac{1}{2}}\underbrace{\Theta^T\Theta}_{I}\Lambda\underbrace{\Theta^T\Theta}_{I}\Lambda^{-\frac{1}{2}} && \text{since } D = D^T \text{ if } D \text{ diagonal} \\
&= \Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}} \\
&= \Lambda(\Lambda^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}) && \text{since diagonal matrices commute} \\
&= \Lambda\Lambda^{-1} \\
&= I
\end{aligned}$$

QED

Special Case: (2x2 matrices)

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

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

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

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

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

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

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

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

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

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

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

Characteristic equation in 2x2 case

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

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

memorize ☆

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

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

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

$T = 4$
$D = 3$

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

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

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

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

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

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

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

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

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

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

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

$= K_{xx}$!

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

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

$$\therefore\ K_{YY} = A K_{XX} A^T$$

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

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

$$= \frac{1}{2}\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.$$

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

$\therefore$ Y_1 and Y_2 are uncorrelated

where $\begin{cases} Y_1 = \frac{1}{\sqrt{6}}(X_1 + X_2) \\ Y_2 = \frac{1}{\sqrt{6}}(X_1 - X_2) \end{cases}$ from $Y = (\Lambda^{-\frac{1}{2}} \vartheta^T) X$

Optimal Estimators

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

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

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

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

Bayesian estimator

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

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

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

and $\Theta \sim h(\theta)$

Maximum Likelihood

$$\hat{\Theta}_n^{ML} = \arg\max_{\theta} g(x \mid \theta) \qquad \theta \in \Theta$$

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

MLE properties

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

2. Asymptotic Normality (from CLT and LLN)

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

I: Fisher Information

3. Invariance Principle (I.P.) ☆

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

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

Log-Likelihood Function L:

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

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

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

Now find <u>MODE</u> 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 $\frac{d^2L}{d\theta^2} < 0$ at $\hat{\theta}_n^{ML}$ value

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

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

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

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

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

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

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

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

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

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

Ex: Normal case.

Find $\underset{2\times1}{\hat{\Theta}_n^{ML}} = \begin{cases} \hat{\mu}_n^{ML} \\ \widehat{(\sigma_n^2)}^{ML} \end{cases}$ if iid $X_1, \dots, X_n \sim N(\mu_x, \sigma_x^2)$

Set $\nabla_\theta L = \Theta$ and solve to give

$$\hat{\Theta}_n^{ML} = \begin{cases} \hat{\mu}_n^{ML} = \bar{X}_n \\ \widehat{(\sigma^2)}^{ML} = \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X}_n)^2 \\ \qquad\quad = \frac{n-1}{n} S_n^2 \end{cases}$$

biased (asymptotically unbiased)

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

$$\widehat{g(\theta)}^{ML} \overset{I.P.}{=} g(\hat{\theta}^{ML})$$

Prf: (only for case when g is 1-1, $u \overset{1-1}{=} g(\theta)$)

$$\therefore \max L(g(\theta)) = \max_{u = g(\theta)} L(u)$$

$$\overset{1-1}{=} \max_u L(g^{-1}(u))$$

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

$\therefore$ Take $\hat{u}^{ML} \overset{1-1}{=} g(\hat{\theta}^{ML})$

QED

Ex: find $\widehat{e^{\cos\theta}}^{ML}$ if iid $X_k \sim \exp(\theta)$

$$\therefore \quad \widehat{e^{\cos\theta}}^{ML} \overset{\text{I.P.}}{=} e^{\cos \hat{\theta}_n^{ML}}$$

$$= e^{\cos \overline{X}_n}$$

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

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