Session 23

Iterated Expectation

23.1

Another common situation:

$$E[g(X,Y)] = \iint_{\mathbb{R}^2} g(x,y)\, f_{X,Y}(x,y)\, dx\, dy$$

$$= \iint_{\mathbb{R}^2} g(x,y)\, f_{Y}(y|x)\, f_X(x)\, dx\, dy$$

$$\int_{\mathbb{R}} f_X(x) \underbrace{\left[\int_{\mathbb{R}} g(x,y)\, f_Y(y|x)\, dy \right]}_{\varphi(x)} dx .$$

$$= \int_{\mathbb{R}} f_X(x)\, E[g(X,Y) \mid X = x]\, dx = \cdots$$

$$(= E[\varphi(X)] .)$$

23.2

$$\cdots = \int_{-\infty}^{\infty} f_X(x)\, \underbrace{E[g(X,Y) \mid X = x]}_{\varphi(x)}\, dx$$

$$= \int_{-\infty}^{\infty} f_X(x)\, \varphi(x)\, dx = E_X[\varphi(X)] \leftarrow$$

where $\varphi(x) = E[g(X,Y) \mid \{X = x\}]$

$$\therefore \quad E[g(X,Y)] = E_X[\varphi(X)]$$

note the notation

$$= E_X[E_Y[g(X,Y) \mid X]]$$

$$= E[E[g(X,Y) \mid X]]$$

22.6

23.3

Summarizing, we have

$$E_{XY}[g(X,Y)] = E_X[E_Y[g(X,Y) \mid X]]$$

<u>n.b.</u> The terminology "iterated" comes from "integrated integration"

$$E[g(X,Y)] = \int_{-\infty}^{\infty} f_X(x) \underbrace{\int_{-\infty}^{\infty} g(x,y)\, f_Y(y \mid \{X=x\})\, dy}_{E_Y[g(X,Y) \mid \{X=x\}]}\, dx$$

One very important application of iterated expectation is <u>Minimum Mean-Square estimation</u>.

So we have

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{XY}(x,y)\, dx\, dy$$

$$= \cdots = \int_{-\infty}^{\infty} f_X(x) \underbrace{\int_{-\infty}^{\infty} g(x,y)\, f_Y(y \mid \{X=x\})\, dy}_{E[g(X,Y) \mid \{X=x\}] = \varphi(x)}\, dx$$

$$= E_X\left[E_Y[g(X,Y) \mid X]\right] = E_X[\varphi(X)].$$

Minimum Mean-Square Estimation

- Let X and Y be two j-dist RVs with j-pdf $f_{XY}(x,y)$.
- Suppose we want to estimate the value of Y given that we have observed $\{X=x\}$.

Q: What is the best estimate of the value of Y given that we know $X=x$?

What do we mean by best?

One commonly used error criterion is square error.

$\Rightarrow$ Design the estimator that minimizes mean-square-error.

23.6

We want to find the function $c(x)$ to estimate the value of $\mathbb{Y}$ given that we observe $\mathbb{X}=x$ suc that

$$\mathcal{E} = E[(\mathbb{Y} - c(\mathbb{X}))^2]$$

is minimized.

Claim: The mean-square error $\mathcal{E}$ is minimized when

$$c(x) = E[\mathbb{Y} \mid \{\mathbb{X}=x\}].$$

23.7

Proof: $\mathcal{E} = E[(\underline{\mathbb{Y}} - c(\underline{\mathbb{X}}))^2]$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - c(x))^2 f_{\mathbb{X}\mathbb{Y}}(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} f_{\mathbb{X}}(x) \underbrace{\left[\int_{-\infty}^{\infty} (y - c(x))^2 \cdot f_{\mathbb{Y}}(y|x)\, dy \right]}_{\varphi(x)} dx$$

$\mathcal{E}$ is minimized if we pick $c(x)$ such that the inner integral is minimized for each value of x.

23.8

We minimize the inner integral as follows (for any particular x):

$$\frac{\partial}{\partial c(x)} \left\{ \int_{-\infty}^{\infty} [y - c(x)]^2 f_{Y}(y|x)\, dy \right\}$$

$$= -2 \int_{-\infty}^{\infty} [y - c(x)] f_{Y}(y|x)\, dy = \underline{0}$$

$$\Rightarrow \int_{-\infty}^{\infty} (y - c(x)) f_{Y}(y|x)\, dy = 0$$

$$\Rightarrow \underbrace{\int_{-\infty}^{\infty} y\, f_{Y}(y|x)\, dy}_{E[Y|X=x]} - c(x) \underbrace{\int_{-\infty}^{\infty} f_{Y}(y|x)\, dy}_{1} =$$

$$E[Y|X=x] - c(x) = 0$$

22.12

23.9

$$\Rightarrow c(x) = E[Y|X=x] \; \star$$

We will use the following notation for the MMSE estimator

$$\hat{Y}_{MMS}(x) = E[Y|X=x]$$

Similarly, by symmetry

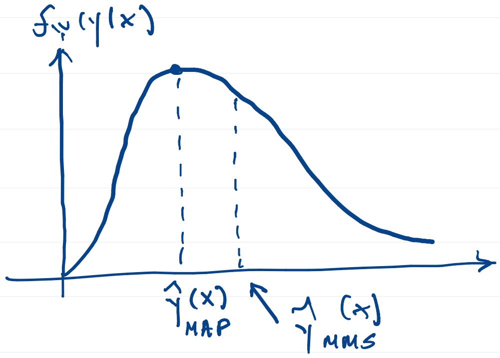
$$\hat{X}_{MMS}(y) = E[X|Y=y] \quad \text{(Exercise)}$$

23.10

Consider another estimator:

$$\hat{Y}_{MAP}(x) = \arg\max_{y} \{ f_Y(y|x) \}$$

MAP $\triangleq$ <u>M</u>aximum <u>A</u>posteriori <u>P</u>robability



The MAP estimators of interest are

$$\hat{Y}_{MAP}(x) = \arg\max_{y} \{ f_Y(y | X = x) \}$$

$$\hat{X}_{MAP}(y) = \arg\max_{x} \{ f_X(x | Y = y) \}.$$

23.11

<u>Papoulis</u> in the reading discussed the Linear Minimum Mean Square Error (LMMSE) estimator

$$c(x) = \underline{ax + b} \;\} \text{ assumed form.}$$

Don't worry about this. You are not responsible for it

Random Vectors

- We've considered two RVs on $(\mathcal{S}, \mathcal{F}, P)$.
- We can extend this to n RVs on $(\mathcal{S}, \mathcal{F}, P)$:

$$\mathbb{X}_1(\omega), \mathbb{X}_2(\omega), \ldots, \mathbb{X}_n(\omega).$$

- We can arrange these RVs as elements of a vector.

(A <u>random vector.</u>)

<u>Defn:</u> Let $\mathbb{X}_1, \ldots, \mathbb{X}_n$ be n jointly distributed RVs on $(\mathcal{S}, \mathcal{F}, P)$. Then the vector of RVs

$$\underline{\mathbb{X}} = (\mathbb{X}_1, \ldots, \mathbb{X}_n) \leftarrow \text{Row Vector}$$

is a <u>random vector</u> (RVec) defined on $(\mathcal{S}, \mathcal{F}, P)$.

Alternatively, we can think of a RVec as a mapping from $\mathcal{S}$ to $\mathbb{R}^n$

$$\underline{\mathbb{X}} : \mathcal{S} \rightarrow \mathbb{R}^n$$

23.14

CDF: $F_{\underline{X}}(\underline{x}) = F_{\underline{X}}(x_1, \dots, x_n)$

$= F_{X_1 \dots X_n}(x_1, \dots, x_n)$

$= P(\{X_1 \le x_1\} \cap \{X_2 \le x_2\} \cap \dots \cap \{X_n \le x_n\})$

PDF: $f_{\underline{X}}(\underline{x}) = f_{\underline{X}}(x_1, \dots, x_n)$

$= f_{X_1 \dots X_n}(x_1, \dots, x_n)$

$$= \frac{\partial^n F_{\underline{X}}(x_1, \dots, x_n)}{\partial x_1 \, \partial x_2 \cdots \partial x_n}$$

23.15

Let $D \subset \mathbb{R}^n$ $(D \in \mathcal{B}(\mathbb{R}^n))$

Then $P(\{\underline{X} \in D\}) = \int_D f_{\underline{X}}(\underline{x})\, d\underline{x}$

$$= \int_{\mathbb{R}^n} f_{\underline{X}}(\underline{x}) \cdot \mathbb{1}_D(\underline{x})\, d\underline{x}$$

$$= \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n\text{-fold integration}} f_{\underline{X}}(x_1, \dots, x_n) \cdot \mathbb{1}_D((x_1, \dots, x_n))\, dx_1 \cdots dx_n$$

23.16

All the general properties of cdfs and pdfs generalize from 2-dimensions to n-dimensions:

e.g. Given j-dist RVs X_1, X_2, X_3, X_4

$$F_{X_1 X_3}(x_1, x_3) = F_{X_1 X_2 X_3 X_4}(x_1, +\infty, x_3, +\infty).$$

or

$$f_{X_1 X_3}(x_1, x_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1 X_2 X_3 X_4}(x_1, x_2, x_3, x_4)\, dx_2\, dx_4$$

27.17

Transformations on RVecs

Given $\underline{X} = (X_1, \dots, X_n) \sim$ RVec

we can define a new RVec

$$\underline{Y} = (Y_1, \dots, Y_k)$$

where $Y_j = g_j(\underline{X})$, $j = 1, \dots, k$

where $k \gtreqless n$

How do we find $F_{\underline{Y}}(\underline{y})$ or $f_{\underline{Y}}(\underline{y})$?

23.18

Given $\underline{X} = (X_1, \dots, X_n) \sim$ RVec

we can define the new RVec

$$\underline{Y} = (Y_1, \dots, Y_k),$$

where

$$Y_j = g_j(\underline{X}), \quad j = 1, \dots, k.$$

Here $k \gtreqless n$.

How do we find $F_{\underline{Y}}(\underline{y})$ or $f_{\underline{Y}}(\underline{y})$?

23.19

To find $F_{\underline{Y}}(y_1, \dots, y_k)$, define

$$D(y_1, \dots, y_k) \triangleq \{(x_1, \dots, x_n) \in \mathbb{R}^n : g_1(x_1, \dots, x_n) \le y_1, \dots, g_k(x_1, \dots, x_n) \le y_k\} \subset \mathbb{R}^n \ (\in \mathcal{B}(\mathbb{R}^n))$$

Then

$$F_{\underline{Y}}(y_1, \dots, y_k) = \int \cdots \int_{D(y_1, \dots, y_k)} f_{\underline{X}}(x_1, \dots, x_n)\, dx_1 \cdots dx_n$$

and

$$f_{\underline{Y}}(y_1, \dots, y_k) = \frac{\partial^k F_{\underline{Y}}(y_1, \dots, y_k)}{\partial y_1\, \partial y_2 \cdots \partial y_k}.$$

Direct Approach (k=n and one-to-one mapping)

Theorem: Given RVec $\underline{X} = (X_1, \ldots, X_n)$, define a new RVec $\underline{Y} = (Y_1, \ldots, Y_n)$ by the one-to-one mapping

$$G : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

described by

$$Y_1 = g_1(\underline{X}), \quad \ldots, \quad Y_n = g_n(\underline{X}).$$

Let $X_1 = h_1(\underline{Y}) = x_1(\underline{Y}), \ldots,$

$X_n = h_n(\underline{Y}) = x_n(\underline{Y}).$

Then the j-pdf of $\underline{Y}$ is

$$f_{\underline{Y}}(y_1, \ldots, y_n) = f_{\underline{X}}(x_1(\underline{y}), \ldots, x_n(\underline{y})) \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} \right|$$

where the Jacobian is given by

$$\frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

<u>n.b.</u> If you have $Y_1, \ldots, Y_k$, where $k < n$, you can introduce aux. variables.