Recap

- $X_1, \cdots X_n$ are continuous rv with joint density

$$Y_1 = g_1(X_1, \cdots, X_n) \quad \cdots \quad Y_n = g_n(X_1, \cdots, X_n)$$

- The transformation is continuous with continuous first partials and is invertible and

$$X_1 = h_1(Y_1, \cdots, Y_n) \quad \cdots \quad X_n = h_n(Y_1, \cdots, Y_n)$$

- We assume the Jacobian of the inverse transform, J, is non-zero
- Then the density of $\mathbf{Y}$ is

$$f_{Y_1 \cdots Y_n}(y_1, \cdots, y_n) = |J| f_{X_1 \cdots X_n}\left(h_1(y_1, \cdots, y_n), \cdots, h_n(y_1, \cdots, y_n)\right)$$

- Called multidimensional change of variable formula

Recap

- One can use the theorem to find densities of sum, difference, product and quotient of random variables.

$$
\begin{aligned}
f_{X+Y}(z) &= \int_{-\infty}^{\infty} f_{XY}(t, z-t)\ dt = \int_{-\infty}^{\infty} f_{XY}(z-t, t)\ dt \\
f_{X-Y}(z) &= \int_{-\infty}^{\infty} f_{XY}(t, t-z)\ dt = \int_{-\infty}^{\infty} f_{XY}(t+z, t) dt
\end{aligned}
$$

$$
f_{X*Y}(z) = \int_{-\infty}^{\infty} \left|\frac{1}{t}\right|\ f_{XY}\left(\frac{z}{t}, t\right)\ dt = \int_{-\infty}^{\infty} \left|\frac{1}{t}\right|\ f_{XY}\left(t, \frac{z}{t}\right)\ dt
$$

$$
f_{(X/Y)}(z) = \int_{-\infty}^{\infty} |t|\ f_{XY}(zt, t)\ dt = \int_{-\infty}^{\infty} \left|\frac{t}{z^2}\right|\ f_{XY}\left(t, \frac{t}{z}\right)\ dt
$$

Recap

- $X_1, X_2, \cdots, X_n$ are said to be exchangeable if their joint distribution is same as that of any permutation of them.
- If the random variables are exchangeable then the joint distribution function remains the same on permutation of arguments.
- Exchangeable random variables are identically distributed but they may not be independent.

Recap

- Let $Z = g(X_1, \cdots X_n) = g(\mathbf{X})$. Then
$$E[Z] = \int_{\Re^n} g(\mathbf{x})\ dF_{\mathbf{X}}(\mathbf{x})$$
- For example, if they have a joint density, then
$$E[Z] = \int_{\Re^n} g(\mathbf{x})\ f_{\mathbf{X}}(\mathbf{x})\ d\mathbf{x}$$
- This gives us: $E[X + Y] = E[X] + E[Y]$
- In general, $E\left[g_1(\mathbf{X}) + g_2(\mathbf{X})\right] = E\left[g_1(\mathbf{X}\right] + E\left[g_2(\mathbf{X})\right]$

- We saw $E[X+Y] = E[X] + E[Y]$.
- Let us calculate $\text{Var}(X+Y)$.

$$\begin{aligned}
\text{Var}(X+Y) &= E\left[((X+Y) - E[X+Y])^2\right] \\
&= E\left[((X - EX) + (Y - EY))^2\right] \\
&= E\left[(X-EX)^2\right] + E\left[(Y-EY)^2\right] \\
&\qquad +2E\left[(X-EX)(Y-EY)\right] \\
&= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)
\end{aligned}$$

where we define **covariance** between X, Y as

$$\text{Cov}(X,Y) = E\left[(X-EX)(Y-EY)\right]$$

- We define **covariance** between X and Y by

$$\begin{aligned}\text{Cov}(X,Y) &= E\left[(X-EX)(Y-EY)\right] \\ &= E\left[XY - X(EY) - Y(EX) + EX\ EY\right] \\ &= E[XY] - EX\ EY\end{aligned}$$

- Note that $\text{Cov}(X,Y)$ can be positive or negative
- X and Y are said to be uncorrelated if $\text{Cov}(X,Y) = 0$
- If X and Y are uncorrelated then

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$$

- Note that $E[X+Y] = E[X] + E[Y]$ for all random variables.

Example

- Consider the joint density

$$f_{XY}(x,y) = 2, \quad 0 < x < y < 1$$

- We want to calculate Cov(X,Y)

$$EX = \int_0^1 \int_x^1 x \; 2 \; dy \; dx = 2 \int_0^1 x \; (1-x) \; dx = \frac{1}{3}$$

$$EY = \int_0^1 \int_0^y y \; 2 \; dx \; dy = 2 \int_0^1 y^2 \; dy = \frac{2}{3}$$

$$E[XY] = \int_0^1 \int_0^y xy \; 2 \; dx \; dy = 2 \int_0^1 y \; \frac{y^2}{2} \; dy = \frac{1}{4}$$

- Hence, Cov$(X,Y) = E[XY] - EX \; EY = \frac{1}{4} - \frac{2}{9} = \frac{1}{36}$

Independent random variables are uncorrelated

- Suppose X, Y are independent. Then

$$
\begin{aligned}
E[XY] &= \int \int x \; y \; f_{XY}(x,y) \; dx \; dy \\
&= \int \int x \; y \; f_X(x) \; f_Y(y) \; dx \; dy \\
&= \int x f_X(x) \; dx \int y f_Y(y) \; dy = EX \; EY
\end{aligned}
$$

- Then, $\mathrm{Cov}(X,Y) = E[XY] - EX \; EY = 0$.
- X, Y independent $\Rightarrow$ X, Y uncorrelated

Uncorrelated random variables may not be independent

- Suppose $X \sim \mathcal{N}(0,1)$ Then, $EX = EX^3 = 0$
- Let $Y = X^2$ Then,

$$E[XY] = EX^3 = 0 = EX\ EY$$

- Thus X, Y are uncorrelated.
- Are they independent? No
  e.g.,
  $$P[X > 2\ |Y < 1] = 0 \neq P[X > 2]$$
- X, Y are uncorrealted does not imply they are independent.

- We define the **correlation coefficient** of X, Y by

$$\rho_{XY} = \frac{\mathsf{Cov}(X, Y)}{\sqrt{\mathsf{Var}(X)\ \mathsf{Var}(Y)}}$$

- If X, Y are uncorrelated then $\rho_{XY} = 0$.
- We will show that $|\rho_{XY}| \leq 1$
- Hence $-1 \leq \rho_{XY} \leq 1,\ \forall X, Y$

- We have $E\left[(\alpha X + \beta Y)^2\right] \geq 0, \ \forall \alpha, \beta \in \Re$

$$
\begin{aligned}
\alpha^2 E[X^2] + \beta^2 E[Y^2] + 2\alpha\beta E[XY] &\geq 0, \quad \forall \alpha, \beta \in \Re \\
\text{Take} \quad \alpha = -\frac{E[XY]}{E[X^2]} & \\
\frac{(E[XY])^2}{E[X^2]} + \beta^2 E[Y^2] - 2\beta \frac{(E[XY])^2}{E[X^2]} &\geq 0, \quad \forall \beta \in \Re \\
\Rightarrow \quad 4\left(\frac{(E[XY])^2}{E[X^2]}\right)^2 - 4E[Y^2]\frac{(E[XY])^2}{E[X^2]} &\leq 0 \\
\Rightarrow (E[XY])^2 \leq E[X^2]E[Y^2] &
\end{aligned}
$$

- We showed that

$$(E[XY])^2 \leq E[X^2]E[Y^2]$$

- Take $X - EX$ in place of X and $Y - EY$ in place of Y in the above algebra.
- This gives us

$$(E[(X - EX)(Y - EY)])^2 \leq E[(X-EX)^2]E[(Y-EY)^2]$$

$$\Rightarrow \quad (\text{Cov}(X,Y))^2 \leq \text{Var}(X)\text{Var}(Y)$$

- Hence we get

$$\rho_{XY}^2 = \left(\frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}\right)^2 \leq 1$$

- The equality holds here only if $E\left[(\alpha X + \beta Y)^2\right] = 0$

Thus, $\quad |\rho_{XY}| = 1$ only if $\alpha X + \beta Y = 0$

- Correlation coefficient of X, Y is ± 1 only when Y is a linear function of X

Linear Least Squares Estimation

- Suppose we want to approximate Y as an affine function of X.
- We want a, b to minimize $E\left[(Y - (aX + b))^2\right]$
- For a fixed a, what is the b that minimizes $E\left[((Y - aX) - b)^2\right]$?
- We know the best b here is: $b = E[Y - aX] = EY - aEX$.
- So, we want to find the best a to minimize $J(a) = E\left[(Y - aX - (EY - aEX))^2\right]$

- We want to find a to minimize

$$\begin{aligned} J(a) &= E\left[(Y - aX - (EY - aEX))^2\right] \\ &= E\left[((Y - EY) - a(X - EX))^2\right] \\ &= E\left[(Y - EY)^2 + a^2(X - EX)^2 - 2a(Y - EY)(X - EX)\right] \\ &= \mathsf{Var}(Y) + a^2\mathsf{Var}(X) - 2a\mathsf{Cov}(X, Y) \end{aligned}$$

- So, the optimal a satisfies

$$2a\mathsf{Var}(X) - 2\mathsf{Cov}(X, Y) = 0 \;\;\Rightarrow\;\; a = \frac{\mathsf{Cov}(X, Y)}{\mathsf{Var}(X)}$$

- The final mean square error, say, J^* is

$$
\begin{aligned}
J^* &= \mathsf{Var}(Y) + a^2\mathsf{Var}(X) - 2a\mathsf{Cov}(X,Y) \\
&= \mathsf{Var}(Y) + \left(\frac{\mathsf{Cov}(X,Y)}{\mathsf{Var}(X)}\right)^2 \mathsf{Var}(X) - 2\frac{\mathsf{Cov}(X,Y)}{\mathsf{Var}(X)} \mathsf{Cov}(X,Y) \\
&= \mathsf{Var}(Y) - \frac{(\mathsf{Cov}(X,Y))^2}{\mathsf{Var}(X)} \\
&= \mathsf{Var}(Y) \left(1 - \frac{(\mathsf{Cov}(X,Y))^2}{\mathsf{Var}(Y)\ \mathsf{Var}(X)}\right) \\
&= \mathsf{Var}(Y) \left(1 - \rho_{XY}^2\right)
\end{aligned}
$$

- The best mean-square approximation of Y as a 'linear' function of X is

$$Y = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}\, X \;+\; \left(EY - \frac{\text{Cov}(X,Y)}{\text{Var}(X)}\, EX \right)$$

- Called the line of regression of Y on X.
- If $\text{cov}(X,Y) = 0$ then this reduces to approximating Y by a constant, EY.
- The final mean square error is

$$\text{Var}(Y)\left(1 - \rho_{XY}^2\right)$$

- If $\rho_{XY} = \pm 1$ then the error is zero
- If $\rho_{XY} = 0$ the final error is $\text{Var}(Y)$

- The covariance of X, Y is

  $$\text{Cov}(X,Y) = E[(X - EX)\ (Y - EY)] = E[XY] - EX\ EY$$

  Note that $\text{Cov}(X,X) = \text{Var}(X)$
- X, Y are called uncorrelated if $\text{Cov}(X,Y) = 0$.
- X, Y independent $\Rightarrow$ X, Y uncorrelated.
- Uncorrelated random variables need not necessarily be independent
- Covariance plays an important role in linear least squares estimation.
- Informally, covariance captures the 'linear dependence' between the two random variables.

Covariance Matrix

- Let $X_1, \cdots, X_n$ be random variables (on the same probability space)
- We represent them as a vector $\mathbf{X}$.
- As a notation, all vectors are column vectors: $\mathbf{X} = (X_1, \cdots, X_n)^T$
- We denote $E[\mathbf{X}] = (EX_1, \cdots, EX_n)^T$
- The $n \times n$ matrix whose $(i,j)^{th}$ element is $\text{Cov}(X_i, X_j)$ is called the covariance matrix (or variance-covariance matrix) of $\mathbf{X}$. Denoted as $\Sigma_{\mathbf{X}}$ or Σ_X

$$\Sigma_{\mathbf{X}} = \begin{bmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \vdots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Cov}(X_n, X_n) \end{bmatrix}$$

Covariance matrix

- If $\mathbf{a} = (a_1, \cdots, a_n)^T$ then
  $\mathbf{a}\,\mathbf{a}^T$ is a $n \times n$ matrix whose $(i,j)^{th}$ element is $a_i a_j$.
- Hence we get

$$\Sigma_{\mathbf{X}} = E\left[(\mathbf{X} - E\mathbf{X})\,(\mathbf{X} - E\mathbf{X})^T\right]$$

- This is because
  $\left((\mathbf{X} - E\mathbf{X})\,(\mathbf{X} - E\mathbf{X})^T\right)_{ij} = (X_i - EX_i)(X_j - EX_j)$
  and $(\Sigma_{\mathbf{X}})_{ij} = E[(X_i - EX_i)(X_j - EX_j)]$

- Recall the following about vectors and matrices
- let $\mathbf{a}, \mathbf{b} \in \Re^n$ be column vectors. Then

$$\left(\mathbf{a}^T\mathbf{b}\right)^2 = \left(\mathbf{a}^T\mathbf{b}\right)^T\left(\mathbf{a}^T\mathbf{b}\right) = \mathbf{b}^T\mathbf{a}\,\mathbf{a}^T\mathbf{b} = \mathbf{b}^T\left(\mathbf{a}\,\mathbf{a}^T\right)\mathbf{b}$$

- Let A be an $n \times n$ matrix with elements a_{ij}. Then

$$\mathbf{b}^T A\mathbf{b} = \sum_{i,j=1}^{n} b_i b_j a_{ij}$$

  where $\mathbf{b} = (b_1, \cdots, b_n)^T$
- A is said to be positive semidefinite if $\mathbf{b}^T A\mathbf{b} \geq 0, \ \forall \mathbf{b}$

- Σ_X is a real symmetric matrix
- It is positive semidefinite.
- Let $\mathbf{a} \in \Re^n$ and let $Y = \mathbf{a}^T \mathbf{X}$.
- Then, $EY = \mathbf{a}^T E\mathbf{X}$. We get variance of Y as

$$\begin{aligned}
\text{Var}(Y) &= E[(Y - EY)^2] = E\left[\left(\mathbf{a}^T\mathbf{X} - \mathbf{a}^T E\mathbf{X}\right)^2\right] \\
&= E\left[\left(\mathbf{a}^T(\mathbf{X} - E\mathbf{X})\right)^2\right] \\
&= E\left[\mathbf{a}^T(\mathbf{X} - E\mathbf{X})\,(\mathbf{X} - E\mathbf{X})^T\mathbf{a}\right] \\
&= \mathbf{a}^T E\left[(\mathbf{X} - E\mathbf{X})\,(\mathbf{X} - E\mathbf{X})^T\right]\mathbf{a} \\
&= \mathbf{a}^T\Sigma_X\mathbf{a}
\end{aligned}$$

- This gives $\mathbf{a}^T\Sigma_X\mathbf{a} \geq 0, \ \forall \mathbf{a}$
- This shows Σ_X is positive semidefinite

- $Y = \mathbf{a}^T \mathbf{X} = \sum_i a_i X_i$ – linear combination of X_i's.
- We know how to find its mean and variance

$$
\begin{aligned}
EY &= \mathbf{a}^T E\mathbf{X} = \sum_i a_i E X_i; \\
\mathsf{Var}(Y) &= \mathbf{a}^T \Sigma_X \mathbf{a} = \sum_{i,j} a_i a_j \mathsf{Cov}(X_i, X_j)
\end{aligned}
$$

- Specifically, by taking all components of $\mathbf{a}$ to be 1, we get

$$
\mathsf{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i,j=1}^{n} \mathsf{Cov}(X_i, X_j) = \sum_{i=1}^{n} \mathsf{Var}(X_i) + \sum_{i=1}^{n} \sum_{j \neq i} \mathsf{Cov}(X_i, X_j)
$$

- If X_i are independent, variance of sum is sum of variances.

- Covariance matrix Σ_X positive semidefinite because

$$\mathbf{a}^T \Sigma_X \, \mathbf{a} = \mathsf{Var}(\mathbf{a}^T \mathbf{X}) \geq 0$$

- Σ_X would be positive definite if $\mathbf{a}^T \Sigma_X \, \mathbf{a} > 0, \; \forall \mathbf{a} \neq 0$
- It would fail to be positive definite if $\mathsf{Var}(\mathbf{a}^T \mathbf{X}) = 0$ for some nonzero $\mathbf{a}$.
- $\mathsf{Var}(Z) = E[(Z - EZ)^2] = 0$ implies $Z = EZ$, a constant.
- Hence, Σ_X fails to be positive definite only if there is a non-zero linear combination of X_i's that is a constant.

- Covariance matrix is a real symmetric positive semidefinite matrix
- It have real and non-negative eigen values.
- It would have n linearly independent eigen vectors.
- These also have some interesting roles.
- We consider one simple example.

- Let $Y = \mathbf{a}^T\mathbf{X}$ and assume $||\mathbf{a}|| = 1$
- Y is projection of $\mathbf{X}$ along the direction $\mathbf{a}$.
- Suppose we want to find a direction along which variance is maximized
- We want to maximize $\mathbf{a}^T\Sigma_X\ \mathbf{a}$ subject to $\mathbf{a}^T\mathbf{a} = 1$
- The lagrangian is $\mathbf{a}^T\Sigma_X\ \mathbf{a} + \eta(1 - \mathbf{a}^T\mathbf{a})$
- Equating the gradient to zero, we get

$$\Sigma_X\ \mathbf{a} = \eta\mathbf{a}$$

- So, $\mathbf{a}$ should be an eigen vector (with eigen value η).
- Then the variance would be $\mathbf{a}^T\Sigma_X\ \mathbf{a} = \eta\mathbf{a}^T\mathbf{a} = \eta$
- Hence the direction is the eigen vector corresponding to the highest eigen value.

Joint moments

- Given two random variables, X, Y
- The joint moment of order (i, j) is defined by

$$m_{ij} = E[X^i Y^j]$$

$m_{10} = EX$, $m_{01} = EY$, $m_{11} = E[XY]$ and so on

- Similarly joint central moments of order (i, j) are defined by

$$s_{ij} = E\left[(X - EX)^i (Y - EY)^j\right]$$

$s_{10} = s_{01} = 0$, $s_{11} = \text{Cov}(X, Y)$, $s_{20} = \text{Var}(X)$ and so on

- We can similarly define joint moments of multiple random variables

- We can define moment generating function of X, Y by

$$M_{XY}(s,t) = E\left[e^{sX+tY}\right], \quad s,t \in \Re$$

- This is easily generalized to n random variables

$$M_{\mathbf{X}}(\mathbf{s}) = E\left[e^{\mathbf{s}^T\mathbf{X}}\right], \quad \mathbf{s} \in \Re^n$$

- Once again, we can get all the moments by differentiating the moment generating function

$$\left.\frac{\partial}{\partial s_i} M_{\mathbf{X}}(\mathbf{s})\right|_{\mathbf{s}=0} = EX_i$$

- More generally

$$\left.\frac{\partial^{m+n}}{\partial s_i^n \, \partial s_j^m} M_{\mathbf{X}}(\mathbf{s})\right|_{\mathbf{s}=0} = EX_i^n X_j^m$$

Conditional Expectation

- Suppose X, Y have a joint density f_{XY}
- Consider the conditional density $f_{X|Y}(x|y)$. This is a density in x for every value of y.
- Since it is a density, we can use it in an expectation integral: $\int g(x)\ f_{X|Y}(x|y)\ dx$
- This is like expectation of $g(X)$ since $f_{X|Y}(x|y)$ is a density in x.
- However, its value would be a function of y.
- That is, this is a kind of expectation that is a function of Y (and hence is a random variable)
- It is called conditional expectation.
- We will now define it formally

- Let X, Y be discrete random variables (on the same probability space).
- The conditinal expectation of $h(X)$ conditioned on Y is a function of Y, and its value for any y is defined by

$$\begin{aligned} E[h(X)|Y = y] &= \sum_x h(x)\ f_{X|Y}(x|y) \\ &= \sum_x h(x)\ P[X = x|Y = y] \end{aligned}$$

- What this means is that we define $E[h(X)|Y] = g(Y)$ where
$$g(y) = \sum_x h(x)\ f_{X|Y}(x|y)$$
- Thus, $E[h(X)|Y]$ is a random variable

- Let X, Y have joint density f_{XY}.
- The conditinal expectation of $h(X)$ conditioned on Y is a function of Y, and its value for any y is defined by

$$E[h(X)|Y = y] = \int_{-\infty}^{\infty} h(x)\ f_{X|Y}(x|y)\ dx$$

- Once again, what this means is that $E[h(X)|Y] = g(Y)$ where

$$g(y) = \int_{-\infty}^{\infty} h(x)\ f_{X|Y}(x|y)\ dx$$

A simple example

- Consider the joint density

$$f_{XY}(x,y) = 2, \;\; 0 < x < y < 1$$

- We calculated the conditional densities earlier

$$f_{X|Y}(x|y) = \frac{1}{y}, \quad f_{Y|X}(y|x) = \frac{1}{1-x}, \;\; 0 < x < y < 1$$

- Now we can calculate the conditional expectation

$$\begin{aligned} E[X|Y=y] &= \int_{-\infty}^{\infty} x \; f_{X|Y}(x|y) \; dx \\ &= \int_0^y x \; \frac{1}{y} \; dx = \frac{1}{y} \; \left. \frac{x^2}{2} \right|_0^y = \frac{y}{2} \end{aligned}$$

- This gives: $E[X|Y] = \frac{Y}{2}$
- We can show $E[Y|X] = \frac{1+X}{2}$

- The conditional expectation is defined by

$$E[h(X)|Y=y] \;=\; \sum_x h(x)\; f_{X|Y}(x|y), \quad X, Y \text{ are discrete}$$

$$E[h(X)|Y=y] \;=\; \int_{-\infty}^{\infty} h(x); f_{X|Y}(x|y)\; dx, \quad X, Y \text{ have joint density}$$

- We can actually define $E[h(X,Y)|Y]$ also as above. That is,

$$E[h(X,Y)|Y=y] = \int_{-\infty}^{\infty} h(x,y)\; f_{X|Y}(x|y)\; dx$$

- It has all the properties of expectation:
  1. $E[a|Y] = a$ where a is a constant
  2. $E[ah_1(X) + bh_2(X)|Y] = aE[h_1(X)|Y] + bE[h_2(X)|Y]$
  3. $h_1(X) \geq h_2(X) \;\;\Rightarrow\;\; E[h_1(X)|Y] \geq E[h_2(X)|Y]$

- Conditional expectation also has some extra properties which are very important
  - $E\left[E[h(X)|Y]\right] = E[h(X)]$
  - $E[h_1(X)h_2(Y)|Y] = h_2(Y)E[h_1(X)|Y]$
  - $E[h(X,Y)|Y = y] = E[h(X,y)|Y = y]$
- We will justify each of these.
- The last property above follows directly from the definition.

- Expectation of a conditional expectation is the unconditional expectation

$$E[\ E[h(X)|Y]\ ] = E[h(X)]$$

In the above, LHS is expectation of a function of Y.
- Let us denote $g(Y) = E[h(X)|Y]$. Then

$$\begin{aligned}
E[\ E[h(X)|Y]\ ] &= E[g(Y)] \\
&= \int_{-\infty}^{\infty} g(y)\ f_Y(y)\ dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} h(x)\ f_{X|Y}(x|y)\ dx \right)\ f_Y(y)\ dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x)\ f_{XY}(x,y)\ dy\ dx \\
&= \int_{-\infty}^{\infty} h(x)\ f_X(x)\ dx \\
&= E[h(X)]
\end{aligned}$$

- Any factor that depends only on the conditioning variable behaves like a constant inside a conditional expectation

$$E[h_1(X)\ h_2(Y)|Y] = h_2(Y)E[h_1(X)|Y]$$

- Let us denote $g(Y) = E[h_1(X)\ h_2(Y)|Y]$

$$\begin{aligned} g(y) &= E[h_1(X)\ h_2(Y)|Y = y] \\ &= \int_{-\infty}^{\infty} h_1(x)h_2(y)\ f_{X|Y}(x|y)\ dx \\ &= h_2(y)\ \int_{-\infty}^{\infty} h_1(x)\ f_{X|Y}(x|y)\ dx \\ &= h_2(y)\ E[h_1(X)|Y = y] \end{aligned}$$

- A very useful property of conditional expectation is $E[\ E[X|Y]\ ] = E[X]$ (Assuming all expectations exist)
- We can see this in our earlier example.

$$f_{XY}(x, y) = 2, \quad 0 < x < y < 1$$

- We calculated: $EX = \frac{1}{3}$ and $EY = \frac{2}{3}$
- We also showed $E[X|Y] = \frac{Y}{2}$

$$E[\ E[X|Y]\ ] = E\left[\frac{Y}{2}\right] = \frac{1}{3} = E[X]$$

- Similarly

$$E[\ E[Y|X]\ ] = E\left[\frac{1+X}{2}\right] = \frac{2}{3} = E[Y]$$

- We have
$$E[\, E[X|Y]\,] = E[X], \quad \forall X, Y$$
- This is a useful technique to find EX.
- We can choose a Y that is useful.