Density of XY

- Let X, Y have joint density f_{XY}.
- Let $Z = XY$. We want to find density of XY directly
- Let $A_z = \{(x, y) \in \Re^2 \ : \ xy \leq z\} \subset \Re^2$.

$$\begin{aligned} F_Z(z) &= P[XY \leq z] = P[(X, Y) \in A_z] \\ &= \int\int_{A_z} f_{XY}(x, y) \, dy \, dx \end{aligned}$$

- We need to find limits for integrating over A_z
- If $x > 0$, then $xy \leq z \;\; \Rightarrow \;\; y \leq z/x$
  If $x < 0$, then $xy \leq z \;\; \Rightarrow \;\; y \geq z/x$

$$F_Z(z) = \int_{-\infty}^{0} \int_{z/x}^{\infty} f_{XY}(x, y) \, dy \, dx + \int_{0}^{\infty} \int_{-\infty}^{z/x} f_{XY}(x, y) \, dy \, dx$$

$$F_Z(z) = \int_{-\infty}^{0} \int_{z/x}^{\infty} f_{XY}(x,y)\; dy\; dx + \int_{0}^{\infty} \int_{-\infty}^{z/x} f_{XY}(x,y)\; dy\; dx$$

- Change variable from y to t using $t = xy$
  $y = t/x;\;\; dy = \frac{1}{x}\, dt;\;\; y = z/x \;\Rightarrow\; t = z$

$$\begin{aligned}
F_Z(z) &= \int_{-\infty}^{0} \int_{z}^{-\infty} \frac{1}{x} f_{XY}(x, \frac{t}{x})\; dt\; dx + \int_{0}^{\infty} \int_{-\infty}^{z} \frac{1}{x} f_{XY}(x, \frac{t}{x})\; dt \\
&= \int_{-\infty}^{0} \int_{-\infty}^{z} \left|\frac{1}{x}\right| f_{XY}(x, \frac{t}{x})\; dt\; dx + \int_{0}^{\infty} \int_{-\infty}^{z} \left|\frac{1}{x}\right| f_{XY}(x, \frac{t}{x}) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z} \left|\frac{1}{x}\right| f_{XY}\left(x, \frac{t}{x}\right)\; dt\; dx \\
&= \int_{-\infty}^{z} \int_{-\infty}^{\infty} \left|\frac{1}{x}\right| f_{XY}\left(x, \frac{t}{x}\right)\; dx\; dt
\end{aligned}$$

This shows: $f_Z(z) = \int_{-\infty}^{\infty} \left|\frac{1}{x}\right| f_{XY}\left(x, \frac{z}{x}\right)\; dx$

Recap: Covariance

- The covariance of X, Y is

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

  Note that $\text{Cov}(X,X) = \text{Var}(X)$
- $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$
- X, Y are called uncorrelated if $\text{Cov}(X,Y) = 0$.
- If X, Y are uncorrelated, $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$
- X, Y independent $\Rightarrow$ X, Y uncorrelated.
- Uncorrelated random variables need not necessarily be independent

Recap: Correlation coefficient

- The correlation coefficient of X, Y is

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

- If X, Y are uncorrelated then $\rho_{XY} = 0$.
- $-1 \leq \rho_{XY} \leq 1, \ \forall X, Y$
- $|\rho_{XY}| = 1$ iff $X = aY$

Recap: mean square estimation

- 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 $\text{Var}(Y)$

Recap: Covariance matrix

- For a random vector, $\mathbf{X} = (X_1, \cdots, X_n)^T$, the covariance matrix is

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

$$(\Sigma_{\mathbf{X}})_{ij} = E[(X_i - EX_i)(X_j - EX_j)]$$

- Var$(\mathbf{a}^T\mathbf{X}) = \mathbf{a}^T\Sigma_X\ \mathbf{a}$
- Σ_X is a real symmetric and positive semidefinite matrix.

Recap: Moment generating function

- For a pair of rv, the joint moment of order (i, j) is $m_{ij} = E[X^i Y^j]$
- The moment generating function of X, Y is $M_{XY}(s,t) = E\left[e^{sX+tY}\right], \;\; s, t \in \Re$
- For n rv, the joint moments are

$$m_{i_1 i_2 \cdots i_n} = E\left[X_1^{i_1} X_2^{i_2} \cdots X_n^{i_n}\right]$$

- The moment generating function of $\mathbf{X}$ is

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

Recap: Conditional Expectation

- The conditional expectation of $h(X)$ conditioned on Y is defined by

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

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

- The conditional expectation of $h(X)$ conditioned on Y is a function of Y: $E[h(X)|Y] = g(Y)$
  the above specify the value of $g(y)$.
- We define $E[h(X,Y)|Y]$ also as above:

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

- If X,Y are independent, $E[h(X)|Y] = E[h(X)]$

Recap: Properties of Conditional Expectation

- It has all the properties of expectation:
  - $E[a|Y] = a$ where a is a constant
  - $E[ah_1(X) + bh_2(X)|Y] = aE[h_1(X)|Y] + bE[h_2(X)|Y]$
  - $h_1(X) \geq h_2(X) \;\; \Rightarrow \;\; E[h_1(X)|Y] \geq E[h_2(X)|Y]$
- Conditional expectation also has some extra properties which are very important
  - $E\left[E[h(X)|Y]\right] = E[h(X)]$
  - $E[h_1(X)h_2(Y)|Y] = h_2(Y)E[h_1(X)|Y]$
  - $E[h(X,Y)|Y = y] = E[h(X,y)|Y = y]$

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

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

Example

- Let X, Y be random variables with joint density given by

$$f_{XY}(x, y) = e^{-y}, \ \ 0 < x < y < \infty$$

- The marginal densities are:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy = \int_{x}^{\infty} e^{-y}\, dy = e^{-x}, \ \ x > 0$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx = \int_{0}^{y} e^{-y}\, dx = y\, e^{-y}, \ \ y > 0$$

  Thus, X is exponential and Y is gamma.

- Hence we have

$$EX = 1; \quad \text{Var}(X) = 1; \quad EY = 2; \quad \text{Var}(Y) = 2$$

$$f_{XY}(x,y) = e^{-y}, \;\; 0 < x < y < \infty$$

- Let us calculate covariance of X and Y

$$\begin{aligned} E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\; f_{XY}(x,y)\; dx\; dy \\ &= \int_{0}^{\infty}\int_{0}^{y} xye^{-y}\; dx\; dy = \int_{0}^{\infty} \frac{1}{2}\; y^3 e^{-y}\; dy = 3 \end{aligned}$$

- Hence, Cov$(X,Y) = E[XY] - EX\; EY = 3 - 2 = 1$.
- $\rho_{XY} = \frac{1}{\sqrt{2}}$

- Recall the joint and marginal densities

$$f_{XY}(x,y) = e^{-y}, \;\; 0 < x < y < \infty$$

$$f_X(x) = e^{-x}, \; x > 0; \quad f_Y(y) = ye^{-y}, \; y > 0$$

- The conditional densities will be

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{e^{-y}}{ye^{-y}} = \frac{1}{y}, \;\; 0 < x < y < \infty$$

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{e^{-y}}{e^{-x}} = e^{-(y-x)}, \;\; 0 < x < y < \infty$$

- The conditional densities are

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

- We can now calculate the conditional expectation

$$E[X|Y = y] = \int x \; f_{X|Y}(x|y) \; dx = \int_0^y x \, \frac{1}{y} \; dx = \frac{y}{2}$$

Thus $E[X|Y] = \frac{Y}{2}$

$$\begin{aligned} E[Y|X = x] &= \int y \; f_{Y|X}(y|x) \; dy = \int_x^\infty y e^{-(y-x)} \; dy \\ &= e^x \left(-y e^{-y}\Big|_x^\infty + \int_x^\infty e^{-y} \; dy \right) \\ &= e^x \left(x e^{-x} + e^{-x} \right) = 1 + x \end{aligned}$$

Thus, $E[Y|X] = 1 + X$

- We got

$$E[X|Y] = \frac{Y}{2}; \quad E[Y|X] = 1 + X$$

- Using this we can verify:

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

$$E[\ E[Y|X]\ ] = E[1 + X] = 1 + 1 = 2 = EY$$

- A property of conditional expectation is

$$E[\ E[X|Y]\ ] = E[X]$$

- We assume that all three expectations exist.
- Very useful in calculating expectations

$$EX = \sum_y E[X|Y = y]\ f_Y(y) \quad \text{or} \quad \int E[X|Y = y]\ f_Y(y)\ dy$$

- Can be used to calculate probabilities of events too

$$P(A) = E[I_A] = E\left[\ E\left[I_A|Y\right]\ \right]$$

- Let X be geometric and we want EX.
- X is number of tosses needed to get head
- Let $Y \in \{0,\ 1\}$ be outcome of first toss. (1 for head)

$$
\begin{aligned}
E[X] &= E[\ E[X|Y]\ ] \\
&= E[X|Y=1]\ P[Y=1] + E[X|Y=0]\ P[Y=0] \\
&= E[X|Y=1]\ p + E[X|Y=0]\ (1-p) \\
&= 1\ p + (1+EX)(1-p) \\
\Rightarrow &\quad EX\ (1-(1-p)) = p + (1-p) = 1 \\
\Rightarrow &\quad EX = \frac{1}{p}
\end{aligned}
$$

- $P[X=k|Y=1]=1$ if $k=1$ (otherwise it is zero) and hence $E[X|Y=1]=1$

$$P[X=k|Y=0] = \begin{cases} 0 & \text{if } k=1 \\ \frac{(1-p)^{k-1}p}{(1-p)} & \text{if } k \geq 2 \end{cases}$$

Hence

$$\begin{aligned} E[X|Y=0] &= \sum_{k=2}^{\infty} k\ (1-p)^{k-2}p \\ &= \sum_{k=2}^{\infty}(k-1)\ (1-p)^{k-2}p + \sum_{k=2}^{\infty}(1-p)^{k-2}p \\ &= \sum_{k'=1}^{\infty} k'\ (1-p)^{k'-1}p + \sum_{k'=1}^{\infty}(1-p)^{k'-1}p \\ &= EX+1 \end{aligned}$$

Another example

- Example: multiple rounds of the party game
- Let R_n denote number of rounds when you start with n people.
- We want $\bar{R}_n = E\left[R_n\right]$.
- We want to use $E\left[R_n\right] = E[\ E\left[R_n | X_n\right]\ ]$
- We need to think of a useful X_n.
- Let X_n be the number of people who got their own hat in the first round with n people.

- R_n – number of rounds when you start with n people.
- X_n – number of people who got their own hat in the first round

$$
\begin{aligned}
E[R_n] &= E[\,E[R_n|X_n]\,] \\
&= \sum_{i=0}^{n} E[R_n|X_n = i]\ P[X_n = i] \\
&= \sum_{i=0}^{n} (1 + E[R_{n-i}])\ P[X_n = i] \\
&= \sum_{i=0}^{n} P[X_n = i] \ + \ \sum_{i=0}^{n} E[R_{n-i}]\ P[X_n = i]
\end{aligned}
$$

If we can guess value of $E[R_n]$ then we can prove it using mathematical induction

- What would be $E[X_n]$?
- Let $Y_i \in \{0,\ 1\}$ denote whether or not i^{th} person got his own hat.
- We know

$$E[Y_i] = P[Y_i = 1] = \frac{(n-1)!}{n!} = \frac{1}{n}$$

$$\text{Now,} \quad X_n = \sum_{i=1}^{n} Y_i \ \text{ and hence } \ EX_n = \sum_{i=1}^{n} E[Y_i] = 1$$

- Hence a good guess is $E[R_n] = n$.
- We verify it using mathematical induction. We know $E[R_1] = 1$

- Assume: $E\left[R_k\right] = k, \;\; 1 \leq k \leq n-1$

$$
\begin{aligned}
E\left[R_n\right] &= \sum_{i=0}^{n} P[X_n = i] \; + \; \sum_{i=0}^{n} E\left[R_{n-i}\right] \; P[X_n = i] \\
&= 1 \; + \; E\left[R_n\right] \; P[X_n = 0] \; + \; \sum_{i=1}^{n} E\left[R_{n-i}\right] \; P[X_n = i] \\
&= 1 \; + \; E\left[R_n\right] \; P[X_n = 0] \; + \; \sum_{i=1}^{n} (n-i) \; P[X_n = i]
\end{aligned}
$$

$$
\begin{aligned}
E\left[R_n\right](1 - P[X_n = 0]) &= 1 + n(1 - P[X_n = 0]) - \sum_{i=1}^{n} i \; P[X_n = i] \\
&= 1 \; + \; n \; (1 - P[X_n = 0]) - E[X_n] \\
&= 1 \; + \; n \; (1 - P[X_n = 0]) - 1
\end{aligned}
$$

$$
\Rightarrow \quad E\left[R_n\right] = n
$$

Analysis of Quicksort

- Given n numbers we want to sort them. Many algorithms.
- Complexity – order of the number of comparisons needed
- Quicksort: Choose a pivot. Separte numbers into two parts – less and greater than pivot, do recursively
- Separating into two parts takes $n - 1$ comparisons.
- Suppose the two parts contain m and $n - m - 1$. Separating both of them into two parts each takes $m + n - m - 1$ comparisons
- So, final number of comparisons depends on the 'number of rounds'

quicksort details

- Given $\{x_1, \cdots, x_n\}$.
- Choose first as pivot

$$\{x_{j_1}, x_{j_2}, \cdots, x_{j_m}\} x_1 \{x_{k_1}, x_{k_2}, \cdots, x_{k_{n-1-m}}\}$$

- Suppose r_n is the number of comparisons. If we get (roughly) equal parts, then

$$r_n \approx n + 2r_{n/2} = n + 2(n/2 + 2r_{n/4}) = n + n + 4r_{n/4} = \cdots = n \log_2(n)$$

- If all the rest go into one part, then

$$r_n = n + r_{n-1} = n + (n-1) + r_{n-2} = \cdots = \frac{n(n+1)}{2}$$

- If you are lucky, $O(n \log(n))$ comparisons.
- If unlucky, in the worst case, $O(n^2)$ comparisons
- Question: 'on the average' how many comparisons?

Average case complexity of quicksort

- Assume pivot is equally likely to be the smallest or second smallest or m^{th} smallest.
- M_n – number of comparisons.
- Define: $X = j$ if pivot is j^{th} smallest
- Given $X = j$ we know $M_n = (n-1) + M_{j-1} + M_{n-j}$.

$$
\begin{aligned}
E[M_n] &= E[\, E[M_n|X] \,] = \sum_{j=1}^{n} E[M_n|X=j]\; P[X=j] \\
&= \sum_{j=1}^{n} E[(n-1) + M_{j-1} + M_{n-j}]\; \frac{1}{n} \\
&= (n-1) + \frac{2}{n} \sum_{k=1}^{n-1} E[M_k], \quad \text{(taking } M_0 = 0\text{)}
\end{aligned}
$$

- This is a recurrence relation. (A little complicated to solve)

Least squares estimation

- We want to estimate Y as a function of X.
- We want an estimate with minimum mean square error.
- We want to solve (the min is over all functions g)

$$\min_{g} \; E\left(Y - g(X)\right)^2$$

- Earlier we considered linear functions: $g(X) = aX + b$
- The solution now turns out to be

$$g^*(X) = E[Y|X]$$

- Let us prove this.

- We want to show that for all g

$$E\left[(E[Y \mid X] - Y)^2\right] \;\leq\; E\left[(g(X) - Y)^2\right]$$

- We have

$$\begin{aligned}
(g(X) - Y)^2 &= \left[(g(X) - E[Y \mid X]) + (E[Y \mid X] - Y)\right]^2 \\
&= \big(g(X) - E[Y \mid X]\big)^2 + \big(E[Y \mid X] - Y\big)^2 \\
&\qquad + 2\big(g(X) - E[Y \mid X]\big)\big(E[Y \mid X] - Y\big)
\end{aligned}$$

- Now we can take expectation on both sides.
- We first show that expectation of last term on RHS above is zero.

First consider the last term

$$
\begin{aligned}
& E\Big[(g(X)-E[Y\mid X])(E[Y\mid X]-Y)\Big] \\
= \;& E\Big[\;\; E\Big\{(g(X)-E[Y\mid X])(E[Y\mid X]-Y)\mid X\Big\}\;\;\Big] \\
& \qquad\qquad \text{because} \quad E[Z]=E[\,E[Z|X]\,] \\
= \;& E\Big[\;\; (g(X)-E[Y\mid X])\;\; E\Big\{(E[Y\mid X]-Y)\mid X\Big\}\;\;\Big] \\
& \qquad\qquad \text{because} \quad E[h_1(X)h_2(Z)|X]=h_1(X)\,E[h_2(Z)|X] \\
= \;& E\Big[\;\; (g(X)-E[Y\mid X])\;\; \Big(E\Big\{(E[Y\mid X])|X\Big\} \;-\; E\{Y\mid X\}\Big)\;\;\Big] \\
= \;& E\Big[\;\; (g(X)-E[Y\mid X])\;\; (E[Y\mid X] \;-\; E[Y\mid X])\;\;\Big] \\
= \;& 0
\end{aligned}
$$

- We earlier got

$$\begin{aligned}(g(X) - Y)^2 \;&=\; \big(g(X) - E[Y \mid X]\big)^2 + \big(E[Y \mid X] - Y\big)^2 \\ &\quad + 2\big(g(X) - E[Y \mid X]\big)\big(E[Y \mid X] - Y\big)\end{aligned}$$

- Hence we get

$$\begin{aligned}E\left[\, (g(X) - Y)^2 \,\right] \;&=\; E\left[\, (g(X) - E[Y \mid X])^2 \,\right] \\ &\quad + E\left[\, (E[Y \mid X] - Y)^2 \,\right] \\ &\geq\; E\left[\, (E[Y \mid X] - Y)^2 \,\right]\end{aligned}$$

- Since the above is true for all functions g, we get

$$g^*(X) \;=\; E\left[Y \mid X\right]$$

Sum of random number of random variables

- Let $X_1, X_2, \cdots$ be iid rv on the same probability space. Suppose $EX_i = \mu, \ \forall i$.
- Let N be a positive integer valued rv that is independent of all X_i.
- Let $S = \sum_{i=1}^{N} X_i$.
- We want to calculate ES. We can use

$$E[S] = E[\ E[S|N]\ ]$$

- We have

$$
\begin{aligned}
E[S|N=n] &= E\left[\sum_{i=1}^{N} X_i \mid N=n\right] \\
&= E\left[\sum_{i=1}^{n} X_i \mid N=n\right] \\
&\qquad \text{since } E[h(X,Y)|Y=y] = E[h(X,y)|Y=y] \\
&= \sum_{i=1}^{n} E[X_i \mid N=n] = \sum_{i=1}^{n} E[X_i] = n\mu
\end{aligned}
$$

- Hence we get

$$E[S|N] = N\mu \quad \Rightarrow \quad E[S] = E[N]E[X_1]$$

- Actually, we did not use independence of X_i.