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

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

- 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: Densities of some standard functions of rv's

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

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

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

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

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: Exchangeable random variables

- $X_1, X_2, \cdots, X_n$ 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: Expectation of function of multiple rv's

- 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

$$E[Z] = \int_{\Re^n} g(\mathbf{x})\; f_{\mathbf{X}}(\mathbf{x})\; d\mathbf{x}$$

- Implies, $E\left[g_1(\mathbf{X}) + g_2(\mathbf{X})\right] = E\left[g_1(\mathbf{X})\right] + E\left[g_2(\mathbf{X})\right]$
- Specifically, $E[X+Y] = E[X] + E[Y]$

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

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

- 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

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

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

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

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

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.

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

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

- We define the characteristic function of $\mathbf{X}$ by

$$\phi_{\mathbf{X}}(\mathbf{s}) = E\left[e^{i\mathbf{s}^T\mathbf{X}}\right], \;\; \mathbf{s} \in \Re^n, \; i = \sqrt{-1}$$

- Characteristic function exists for all random vectors and it uniquely determines the joint distribution.
- The moment generating function may not exist for all random vectors. When it exists, it uniquely determines the joint distribution.

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.

- 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

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

- That is, we define $E[h(X)|Y] = g(Y)$ where

$$E[h(X)|Y=y] = g(y) = \sum_x h(x)\ f_{X|Y}(x|y)$$

- 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

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

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

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

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

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

- 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.

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

- 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

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

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

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

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

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)

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

- $P[X = k|Y = 1] = 1$ 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

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

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

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

Now, $\quad X_n = \sum_{i=1}^{n} Y_i$ 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: $E[R_k] = k, \;\; 1 \leq k \leq n-1$

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

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

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

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 $n - m - 1$. Comparisons needed to Separate each of them into two parts depends on m
- 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 only linear functions: $g(X) = aX + b$
- Now we want the 'best' function (linear or nonlinear)
- 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

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

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

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