Random Variables and Expectation

Curs 2019

Random variables

Flip 5 times a fair coin, each time if the outcome is H we give 1€, if it is T we get 1€. At the end, how much did we win or loose?. Notice $\Omega = \{T, H\}^5$

Given Ω, a random variable is a function $X : \Omega \to \mathbb{R}$.
X can be interpreted as a quantity, whose value depends on the outcome of the experiment.

For ex. in the previous example, define X to be the number wins-looses $(\#T - \#H)$.
In the above example $\mathbf{Pr}[X = -1] = 5/16$ and $\mathbf{Pr}[X = 2] = 0$

Events $\leftrightarrow$ random variables

$\leftarrow$ Given a random variable X on Ω and $a \in \mathbb{R}$ the event $X \geq a$ represents the set $\{\omega \in \Omega | X(\omega) \geq a\}$.

$$\mathbf{Pr}[X \geq a] = \sum_{\omega \in \Omega : X(\omega) \geq a} \mathbf{Pr}[\omega] .$$

Similar events could be associated with $X \leq a$, $X = a$, ...

For ex. in the previous example, for $(X = 1)$, we can associate with the event A that we get exactly 1€, (i,e, we get 3 T and 2 H), and we have

$$\mathbf{Pr}[X = 1] = \mathbf{Pr}[A] = 1 = \frac{\binom{5}{3}}{2^5} .$$

Trick 2: Indicator random variables

$\Rightarrow$ Given an event A define the indicator random variable I_A:

$$I_A = \begin{cases} 1 & \text{if } A \text{ true} \\ 0 & \text{otherwise} \end{cases}$$

For ex. if $A =$ the event of wining euro, then
$\mathbf{Pr}[A] = \mathbf{Pr}[I_A = 1] = \frac{\binom{5}{3}}{2^5}$.

An indicator random variable of an event gives the probability that the event happens.

Expectation

The expectation of a r.v. $X : \Omega \to \mathbb{R}$, $\mathbf{E}[X]$ is defined as

$$\mathbf{E}[X] = \sum_{x \in X(\Omega) \subset \mathbb{R}} x \cdot \mathbf{Pr}[X = x] = \sum_{\omega \in \Omega} X(\omega) \mathbf{Pr}[\omega].$$

Define the rang(X), as $X(\Omega) \subset \mathbb{R}$.

Expectation (mean, average) is just the weighted sum over all values of the r.v.

Notice: If X is a r.v. then $\mathbf{E}[X] \in \mathbb{R}$.

Let X be an integer generated u.a.r. between 1 and 6 ($X =$Rand$(1,6)$). Then $\mathbf{E}[X] = \sum_{x=1}^{6} x \cdot \mathbf{Pr}[X = x] = \sum_{x=1}^{6} \frac{x}{6} = 3.5$.

Notice that in the previous example, $\mathbf{E}[X]$ is not one of the possible values for X.

Example

Define the r.v. $X = \text{Rand}(1,5) + \text{Rand}(1,6)$. Compute $\mathbf{E}[X]$.

$\Omega = \{(i,j) | (1 \le i \le 5) \wedge (1 \le j \le 6)\}$, therefore $|\Omega| = 30$,

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{i=2}^{11} i \mathbf{Pr}[X = i] \\
&= \frac{1}{30}2 + \frac{2}{30}3 + \frac{3}{30}4 + \frac{4}{30}5 + \frac{5}{30}6 + \\
&+ \frac{5}{30}7 + \frac{4}{30}8 + \frac{3}{30}9 + \frac{2}{30}10 + \frac{1}{30}11 = 6.5
\end{aligned}$$

Trick 3: Linearity of expectation

Theorem Given r.v. X, Y on Ω and a $c \in \mathbb{R}$,

- $\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$.
- Given $c \in \mathbb{R}$, and a rv X, then $\mathbf{E}[cX] = c\mathbf{E}[X]$.

Proof: • Let $Z = X + Y$,

$$\begin{aligned}
\mathbf{E}[Z] &= \sum_{\omega\in\Omega} \mathbf{Pr}[\omega]\, Z(\omega) = \sum_{\omega\in\Omega} \mathbf{Pr}[\omega]\,(X(\omega) + Y(\omega)) \\
&= \sum_{\omega\in\Omega} (\mathbf{Pr}[\omega]\, X(\omega) + \mathbf{Pr}[\omega]\, Y(\omega)) \\
&= \sum_{\omega\in\Omega} \mathbf{Pr}[\omega]\, X(\omega) + \sum_{\omega\in\Omega} \mathbf{Pr}[\omega]\, Y(\omega) = \mathbf{E}[X] + \mathbf{E}[Y]
\end{aligned}$$

• $\mathbf{E}[cX] = \sum_{x\in\mathbb{R}} cx \cdot \mathbf{Pr}[X = x] = c\sum_{x\in\mathbb{R}} x \cdot \mathbf{Pr}[X = x]$

Corollary Given r.v. $\{X_i\}_{i=1}^n$, $\mathbf{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbf{E}[X_i]$.

Trick 2b: Expectation of an Indicator r.v.

Recall: If A is the event and I_A its indicator random variable, then

$$\mathbf{E}[A] = \mathbf{Pr}[I_A].$$

Proof $\mathbf{E}[I_A] = \sum_{\omega \in A} 1 \cdot \mathbf{Pr}[\omega] + \sum_{\omega \notin A} 0 \cdot \mathbf{Pr}[\omega] = \mathbf{Pr}[A].$ □

Ex.: A professor gave a text to n students and after u.a.r. returns each test to a random student, so each student grades the test s/he gets. What is the expected number of students grading their own test?

The probability that any student gets its own test is $1/n$. Let X_i an indicator r.v. such that is 1 if the ith student gets its own test. Let $X = \sum_{i=1}^{n} X_i$. Then

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

Jensen's inequality

Recall f is convex in $[a, b]$ if $\forall x_1, x_2 \in [a, b]$ and $\forall t \in [0, 1]$ we have $f[tx_1 + (1-t)x_2] \leq tf(x_1) + (1-t)f(x_2)$.

Lemma If f is convex then $\mathbf{E}[f(X)] \geq f(\mathbf{E}[X])$.
Proof Let $\mu = \mathbf{E}[X]$ $(\mu \in \mathbb{R})$. Using Taylor to expand f at $X = \mu$,

$$\begin{aligned} f(X) &= f(\mu) + f'(\mu)(X-\mu) + \frac{f''(\mu)(X-\mu)^2}{2} + \cdots \\ &\geq f(\mu) + f'(\mu)(X-\mu) \\ \mathbf{E}[f(X)] &\geq \mathbf{E}\left[f(\mu) + f'(\mu)(X-\mu)\right] \\ &= \mathbf{E}[f(\mu)] + f'(\mu)(\mathbf{E}[X] - \mu) = f(\mu) \qquad \square \end{aligned}$$

i.e $\mathbf{E}[f(X)] \geq f(\mathbf{E}[X])$.

Some applications of Jensen's inequality

- $\mathbf{E}\left[X^2\right] \geq (\mathbf{E}[X])^2$;
- $\mathbf{E}\left[e^X\right] \geq e^{\mathbf{E}[X]}$;
- $\mathbf{E}[X \ln X] \geq \mathbf{E}[X] \ln \mathbf{E}[X]$, if $X \in (0, \infty)$.

Probability Mass Function

Let X have as value an integer chosen u.a.r. in $[1, 15]$.

To compute $\mathbf{E}[X \mod 5]$

Define a new rv $Y = X \mod 5$, the rang $X(\Omega)$ of $Y = \{0, 1, 2, 3, 4\}$

$$\begin{aligned}\mathbf{E}[Y] &= 0\mathbf{Pr}[Y=0] + 1\mathbf{Pr}[Y=1] + 2\mathbf{Pr}[Y=2] + 3\mathbf{Pr}[Y=3] + 4\mathbf{Pr}[Y=4] \\ &= \mathbf{Pr}[\{1,6,11\}] + 2\mathbf{Pr}[\{2,7,12\}] + 3\mathbf{Pr}[\{3,8,13\}] + 4\mathbf{Pr}[\{4,8,14\}] \\ &= \frac{3}{15} + 2\frac{3}{15} + 3\frac{3}{15} + 4\frac{3}{15} = 1.8\end{aligned}$$

Notice the only thing needed to compute $\mathbf{E}[Y]$ is $\mathbf{Pr}[Y = i]$, for $i \in [1, 15]$.

Probability Mass Function

Let X be a discrete random variable, then its range (possible values for $X(\Omega)$) is a countable set.
Given a r.v X, define its probability mass function (PMF) as $p_X : \mathbb{R} \to \mathbb{R}$ s.t. $\forall x \in X(\Omega), p_X(x) = \mathbf{Pr}\left[X = x\right]$.

i.e. Given X the PDF defines the $\mathbf{Pr}\left[X = x\right]$ for each x in the range of X.

Ex., Let X take u.a.r. integer values in $[1, \dots, 10]$, what is $\mathbf{E}\left[X \mod 3\right]$?
$\Omega = \{1, \dots, 10\}$. Let $Y = X \mod 3$. The range of $Y = \{0, 1, 2\}$
$\mathbf{E}\left[Y\right] = 0\mathbf{Pr}\left[Y = 0\right] + 1\mathbf{Pr}\left[Y = 1\right] + 2\mathbf{Pr}\left[Y = 2\right] =$
$\mathbf{Pr}\left[Y = 1\right] + 2\mathbf{Pr}\left[Y = 2\right] = 4/10 + 2(3/10) = 1$

PMF

Knowing the $p_X()$ of a r.v. X, it means to know the range $X(\Omega) = \{x_i\}_i$,
Moreover, for every $x_i \in X(\Omega)$, we have the value $p_X(x_i)$.

Note that a PMF $= 0$ on most inputs x, it's only nonzero on $x \in X(\Omega)$.

The definition of the expected value can be rewritten as

$$\mathbf{E}[X] = \sum_{x \in X(\Omega)} x \cdot p_X(x)$$

A PMF $p_X(x)$ has the following properties:

- $p_X(x) \geq 0$,
- $\sum_{x \in X(\Omega)} p_X(x) = 1$,
- for any $S \subseteq \mathbb{R}$, $\mathbf{Pr}[X \in S] = \sum_{x \in S} p_X(x)$.

Useful property of PMF

If X is a discrete r.v. with with PMF p_X and g is a real-valued function then $\mathbf{E}\left[g(X)\right] = \sum_{x_i} g(x_i)p_X(x_i)$.

Example: Let X with $p_X(-1) = 0.2, p_X(0) = 0.5, p_X(1) = 0.3$.
Let $Y = g(X) = X^2$.
Then, $\mathbf{E}\left[Y\right] = (-1)^2 0.2 + 0^2 0.5 + 1^2 0.3 = 0.5$.

Joint PMF

Consider the following experiment:
$X = \text{RandI}(1,2)$ and $Y = \text{RandI}(1, X+1)$ (Y depends on X)

$\Omega = \{(1,1),(1,2),(2,1),(2,2),(2,3)\}$
Therefore, $p_X(1) = 1/2$; $p_X(2) = 1/2$ and
$p_Y(1) = 1/4 + 1/6 = 5/12$; $p_Y(2) = 5/12$; $p_Y(3) = 1/6$.

$$\begin{aligned}\mathbf{E}[XY] &= \sum_{\omega \in \Omega} \mathbf{Pr}[\omega] X(\omega) Y(\omega) \\ &= \frac{1}{4} \cdot 1 \cdot 1 + \frac{1}{4} \cdot 1 \cdot 2 + \frac{1}{6} \cdot 2 \cdot 1 + \frac{1}{6} \cdot 2 \cdot 2 + \frac{1}{6} \cdot 2 \cdot 3 = \frac{11}{4}.\end{aligned}$$

BUT $\mathbf{E}[X] = 3/2$ and $\mathbf{E}[Y] = 7/4$ so $\mathbf{E}[X]\,\mathbf{E}[Y] = 21/8$.
Therefore,

$$\mathbf{E}[XY] \neq \mathbf{E}[X]\,\mathbf{E}[Y].$$

Joint PMF

A said, in this example X and Y are not independent, and $\mathbf{E}[XY] = \frac{11}{4} \neq \frac{21}{8} = \mathbf{E}[X]\,\mathbf{E}[Y]$.

Usually the joint PMF of 2 or more r.v. is given by a table:

		Y		
		1	2	3
X	1	1/4	1/4	0
	2	1/6	1/6	1/6

Given a joint PMF of X and Y, it's easy to get the PMF of each of them,

- To get the PMF of X sum the rows,
- to get the PMF of Y sum the columns.

Joint PMF

The joint PMF of r.v. X, Y is the function $p_{XY} : \mathbb{R}^2 \to \mathbb{R}$ defined by $p_{XY}(x, y) = \mathbf{Pr}[X = x \wedge Y = y]$.

Useful equation: With the joint PMF of r.v. X, Y you can compute the expectation of any function $f(X, Y)$:

$$\mathbf{E}[f(X, Y)] = \sum_{x,y} p_{XY}(x, y) \cdot f(x, y).$$

Compute $\mathbf{E}\left[\frac{X}{Y}\right]$ for the previous r.v. X, Y

$$\begin{aligned} \mathbf{E}[f(X, Y)] = {} & p_{XY}(1,1)\frac{1}{1} + p_{XY}(1,2)\frac{2}{1} + p_{XY}(1,3)\frac{3}{1} \\ & + p_{XY}(2,1)\frac{1}{2} + p_{XY}(2,2)\frac{2}{2} + p_{XY}(3,2)\frac{2}{3} = \frac{5}{4} \end{aligned}$$

Independent r.v.

Two random variables X and Y are said to be independent if

$$\forall x, y \in \mathbb{R}, \mathbf{Pr}\left[(X = x) \cap (Y = y)\right] = \mathbf{Pr}\left[X = x\right] \cdot \mathbf{Pr}\left[Y = y\right].$$

Two not independent r.v. are said to be correlated.

Rolling two dices, let X_1 be a r.v. counting the pips in dice 1, and let X_2 be a r.v. counting the pips in dice 2. Then X_1 and X_2 are independents.
Rolling two dices, let X_1 be a r.v. counting the pips in dice 1, and let X_3 count the sum of pips in the two rollings, then X_1 and X_3 are correlated.

Independent r.v.: Main result

Theorem If X and Y are independent r.v. then $\mathbf{E}[XY] = \mathbf{E}[X]\,\mathbf{E}[Y]$.

Proof

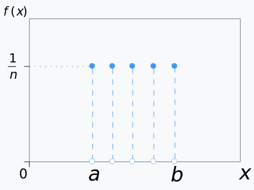
$$\begin{aligned}
\mathbf{E}[XY] &= \sum_{x,y} p_{XY}(x,y) \cdot x \cdot y \\
&= \sum_{x,y} p_X(x) p_Y(y) \cdot xy \text{ (by independence)} \\
&= \sum_{x,y} x p_X(x) y p_Y(y) = \sum_{x} x p_X(x) \sum_{y} y p_Y(y) \\
&= \mathbf{E}[X]\,\mathbf{E}[Y] \qquad \square
\end{aligned}$$

Recall that if X and Y are independent, then for any real value f and g, $f(X)$ and $g(Y)$ also are independent
$\Rightarrow \mathbf{E}[f(X) \cdot g(Y)] = \mathbf{E}[f(X)] \cdot \mathbf{E}[g(Y)]$

Probability Distribution: The uniform

We have been using it: A discrete r.v. X, with values $x_1, \ldots, x_n$ follows a uniform distribution $X \in \mathcal{U}$ if any x_i is selected with equal probability $\mathbf{Pr}[X = x_i] = 1/n$.

$$\mathbf{E}[X] = \sum_{i=1}^{n} x_i \mathbf{Pr}[X = x_i] = \frac{\sum_{i=1}^{n} x_i}{n}.$$



Bernoulli Process

A Bernoulli process denotes a sequence of experiments, each of them a with binary output: success (1 or T) with probability p, and failure (0 ot F) with prob. $q = 1 - p$.

Therefore $\Omega = \{0, 1\}^n$ for $n \geq 1$

Nice thing about Bernoulli distributions: it is natural to use indicator random variable.

The two classical distributions associated with Bernoulli processes are the binomial and the geometric.

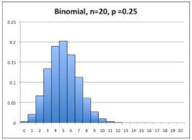
The Binomial distribution

Flip a coin n times, which is the probability of having exactly k heads?

A r.v. X has a Binomial distribution with parameter p $(X \in B(n,p))$ if X counts the number of occurrences during n independent trials, each with $\mathbf{Pr}[\text{success}] = p$.

The PMF:

$$\mathbf{Pr}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$



If $X \in B(n,p)$, then $\mathbf{E}[X] = np$.

Proof To compute $\mathbf{E}[X]$, we define indicator r.v. $\{X_i\}_{i=1}^n$, where $X_i = 1$ iff i-th. output is 1, otherwise $X_i = 0$.

Then, $X = \sum_{i=1}^n X_i \Rightarrow \mathbf{E}[X] = \mathbf{E}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \underbrace{\mathbf{E}[X_i]}_{=p} = np.$

$\square$

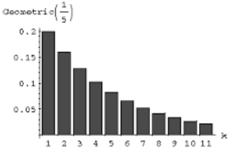
The Geometric distribution

How many flip a coin, until getting a H?

A r.v. X has a Geometric distribution with parameter p ($X \sim G(p)$) if X counts the number of trials until the first success.

Notice that if $X \sim G(p)$, $\Omega = \{0,1\}^k$, for $k \geq 1$ and $X(\Omega) = \{1,2,3,4,\dots\}$

If $X \in G(p)$ then its PMF:
$\mathbf{Pr}[X = k] = (1-p)^{k-1}p$,



Recall: If $x < 1$ the Geometric series is $\sum_{k=0}^{\infty} x^k = \frac{1}{(1-x)}$, then

$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=1}^{\infty}(1-p)^{k-1}p = \frac{p}{1-(1-p)} = 1$$

The Geometric distribution

Therorem If $X \in G(p)$ then $\mathbf{E}[X] = \frac{1}{p}$.

Proof. Let $q = (1 - p)$

$\mathbf{E}[X] = \sum_{k=1}^{\infty} k\mathbf{Pr}[X = k] = \sum_{k=1}^{\infty} kq^{k-1}p = p+2qp+3q^2p+4q^3p+\cdots$

Trick to evaluate the infinite series: multiply by q

$$q\mathbf{E}[X] = qp + 2q^2p + 3q^3p + 4q^4p \ldots$$

$$\mathbf{E}[X] - (1-p)\mathbf{E}[X] = \underbrace{p + qp + q^2p + q^3p + \cdots}_{=1}$$

Therefore, $\mathbf{E}[X] = \frac{1}{p}$. $\square$

Geometric Distribution: Memoryless property

Memoryless property: The future is independent of the past

Flipping a fair coin the past sequence HTHTTTTTTTTTTT?

doesn't tell us what next outcome ? would be.

Lemma If $X \in G(p)$, for any $n > 0$,

$$\mathbf{Pr}\left[X = n + k \mid X > k\right] = \mathbf{Pr}\left[X = n\right].$$

Proof. Recall for $0 < x < 1$ $\sum_{i=k}^{\infty} x^i = x^k/(1-x)$.

$$\begin{aligned}
\mathbf{Pr}\left[X = n + k \mid X > k\right] &= \frac{\mathbf{Pr}\left[(X = n + k) \cap (X > k)\right]}{\mathbf{Pr}\left[X > k\right]} \\
&= \frac{\mathbf{Pr}\left[X = n + k\right]}{\mathbf{Pr}\left[X > k\right]} = \frac{q^{n+k-1}p}{\sum_{i=k}^{\infty} q^i p} \\
&= \frac{q^{n+k-1}p}{q^k} = q^{n-1}p = \mathbf{Pr}\left[X = n\right] \quad \square
\end{aligned}$$

We used the equation: $\mathbf{Pr}\left[X \geq n\right] = \sum_{k \geq n} q^{k-1} p = q^{n-1}$.

The Poisson approximation to the Binomial

For $X \in B(n, p)$, for large n, computing the PMF $\mathbf{Pr}[X = x]$ could be quite nasty.
It turns out that for large n and small p, $B(n, p)$ can be easily approximated by the PMF of a simpler Poisson random variable.

A discrete r.v. X is Poisson with parameter λ ($X \in P(\lambda)$), if it has PMF $\mathbf{Pr}[X = i] = \frac{\lambda^i e^{-\lambda}}{i!}$, for $i \in \{0, 1, 2, 3, \ldots\}$

If $X \in P(\lambda)$ then $\mathbf{E}[X] = \lambda$.

This is the reason that sometimes λ is denoted μ.

Proof:

$\mathbf{E}[X] = \sum_{i=1}^{\infty} i \frac{\lambda^i e^{-\lambda}}{i!} = e^{-\lambda} \lambda \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = e^{-\lambda} \lambda e^{\lambda} = \lambda$ $\square$

The Poisson approximation to the Binomial

Theorem: If $X \in B(n, p)$, with $\mu = pn$, then as $n \to \infty$, for each fixed $i \in \{0, 1, 2, 3, \ldots\}$,

$$\mathbf{Pr}[X = i] \sim \frac{\mu^i e^{-\mu}}{(i-1)!}.$$

Proof: As $\mu = np$,

$$\begin{aligned}
\mathbf{Pr}[X = i] &= \binom{n}{i} (\frac{\mu}{n})^i (1 - \frac{\mu}{n})^{n-i} \\
&= \frac{n(n-1)\cdots(n-i+1)}{i!} \frac{\mu^i}{n^i} (1 - \frac{\mu}{n})^n (1 - \frac{\mu}{n})^{-i} \\
&= \frac{\mu^i}{i!} (1 - \frac{\mu}{n})^n \frac{n(n-1)\cdots(n-i+1)}{n^i} (1 - \frac{\mu}{n})^{-i} \\
&\sim \frac{\mu^i}{i!} e^{-\mu} \text{ as } n \to \infty. \qquad \square
\end{aligned}$$

Example

The population of Catalonia is around 7 million people. Assume Suppose that the probability that a person is killed by lightning in a year is, independently, $p = \frac{1}{5\times 10^8}$.

a.- Compute the exact probability that 3 or more people will be killed by lightning next year in Catalonia.

Let X be a r.v. counting the number of people that will be killed in Cat. next year by a lightning.

We want to compute

$\mathbf{Pr}[X \geq 3] = 1 - \mathbf{Pr}[X \geq 0] - \mathbf{Pr}[X = 1] - \mathbf{Pr}[X = 2]$, where $X \in B(7 \times 10^6, \frac{1}{5\times 10^8})$.

Then,

$\mathbf{Pr}[X \geq 3] = 1 - (1-p)^n - np(1-p)^{n-1} - \binom{n}{2} p^2 (1-p)^{n-2} = 1.65422 \times 10^{-7}$

Example

b.- Approximate $\mathbf{Pr}[X \geq 3]$ $\lambda = np = 7/500$ so
$\mathbf{Pr}[X \geq 3] \sim 1 - e^{\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2}{2} e^{-\lambda} = 1.52558 \times 10^{-7}$

c.- Approximate the probability that 2 or more people will be killed by lightning the first 6 months of 2019
Notice we are considering λ as a *rate*. Then $\lambda = 7/2 \times 500$
$\mathbf{Pr}[X \geq 2 \text{ during 6 months}] \sim 1 - e^{\lambda} - \lambda e^{-\lambda} = 5.79086 \times 10^{-7}$

d.- Approximate the probability that in 3 of the next 10 years exactly 3 people will be killed
We have $\lambda = 7/500$, then the probability that every year 3 people are killed $= \frac{e^{-\lambda}\lambda^3}{3!}$. Let Y be a r.v. counting the number of years with exactly 3 kills.
Assuming independence between years, $Y \in (19, \frac{e^{-\lambda}\lambda^3}{3!})$, therefore the answer is $\binom{10}{3}(\frac{e^{-\lambda}\lambda^3}{3!})^3(1 - \frac{e^{-\lambda}\lambda^3}{3!})^7$

Conditional Expectation

Let X be a r.v. and A be an event, define the conditional expectation of X given E:

$$\mathbf{E}\left[X|A\right] = \sum_{x \in X(\Omega)} x\mathbf{Pr}\left[X = x|A\right].$$

In general the event A will be of the form $Y = y$ for a r.v. Y on Ω, so we get an alternative definition of conditional expectation: $\mathbf{E}\left[X|Y = y\right] = \sum_{x \in X(\Omega)} x\mathbf{Pr}\left[X = x|Y = y\right]$.

Important: For given r.v. X and Y, $\mathbf{E}\left[X\right]$ is a constant real, while $\mathbf{E}\left[X|Y\right]$ is itself a random variable, taking value $\mathbf{E}\left[X|Y = y\right]$, when $Y = y$.

Interesting Example

Roll two dices. Let X be a r.v. counting the sum of the picks, and let X_1 be the r.v. counting the number of picks in the first dice. We want: (1.-) $\mathbf{E}[X_1|X=5]$, and (2.-) $\mathbf{E}[X|X_1]$.

(1.-) $\mathbf{E}[X_1|X=5] = \sum_{x=1}^{4} x\mathbf{Pr}[X_1 = x|X=5]$

But $\mathbf{Pr}[X_1 = x|X=5] = \frac{\mathbf{Pr}[(X_1=x)\cap(X=5)]}{\mathbf{Pr}[X=5]} = \frac{1/36}{4/36} = \frac{1}{4}$

Therefore, $\mathbf{E}[X_1|X=5] = \frac{1+2+3+4}{4} = \frac{5}{2}$

(2.-)
$\mathbf{E}[X|X_1] = \sum_{2\leq x\leq 12} x\mathbf{Pr}[X=x|X_1] = \sum_{x=X_1+1}^{X_1+6} \frac{x}{6} = X_1 + \frac{7}{6}$.

So $\mathbf{E}[X|X_1]$ is a r.v. whose value depends on X_1

$\therefore \mathbf{E}[\mathbf{E}[X|X_1]] = \mathbf{E}\left[X_1 + \frac{7}{2}\right] = \frac{7}{2} + \frac{7}{2} = 7 = \mathbf{E}[X]$.

Formalization of the previous example

Lemma: Let X, Y be r.v., then

$$\mathbf{E}[X] = \sum_{y \in Y(\Omega)} \mathbf{Pr}[Y = y]\, \mathbf{E}[X | Y = y].$$

Proof:

$$\begin{aligned}
\sum_y \mathbf{Pr}[Y = y]\, \mathbf{E}[X | Y = y] &= \sum_y \mathbf{Pr}[Y = y] \sum_x x \mathbf{E}[X = x | Y = y] \\
&= \sum_x \sum_y x \mathbf{E}[X = x | Y = y]\, \mathbf{Pr}[Y = y] \\
&= \sum_x \sum_y x \mathbf{Pr}[(X = x) \cap (Y = y)] \\
&= \sum_x x \mathbf{Pr}[X = x] = \mathbf{E}[X] \qquad \square
\end{aligned}$$

Formalization of the previous example

Recall we have seeing that for any r.v. Z, Y, then $\mathbf{E}[Y|Z]$ is just a r.v. $f(Z)$ s.t. $\mathbf{E}[Y|Z] = \mathbf{E}[Y|Z = z]$, when $Z = z$.

Theorem: Let X, Y, Z be r.v. on Ω. Then,

$$\mathbf{E}[Y] = \mathbf{E}[\mathbf{E}[Y|Z]]$$

. **Proof:** Using the previous lemma,
$\mathbf{E}[\mathbf{E}[Y|Z]] = \sum_z \mathbf{E}[Y|Z = z] \mathbf{Pr}[Z = z] = \mathbf{E}[Y]$ $\square$

Deterministic algorithm to hire a student

We have n students $\{1, \ldots, n\}$, we want to hire the best one to help us. For that we have to interview one by one, each time we find one that is more suitable that the previous ones, we preselect him. At the end we hire the last one pre-selected, but we indemnify with $S > 0$ €, each of the pre-selected not hired. We want to minimize the number of students pre-selected.

Hiring (n)
best:=0
for $i = 1$ to n **do**
  interview i
  if i is better than best **then**
    best:=i and pre-select i
  end if
end for

Adversarial complexity

The adversary gives you a list of ordered students s.t. you are forced to pre-select each of them.

$$T(n) = \Theta(n).$$

Average analysis of the hiring algorithm

The number of all possible order of the students is $n!$
We select u.a.r. an order with probability $= \frac{1}{n!}$.

Lemma The expected number of pre-selected is $O(\lg n)$
Proof Let X be a r.v. counting the number of pre-selected students.
For each $1 \leq i \leq n$ definean indicator r.v.

$$X_i = \begin{cases} 1 & \text{if } i \text{ is-pre-selected;} \\ 0 & \text{otherwise.} \end{cases}$$

Then,
$$X = \sum_{i=1}^{n} X_i \Rightarrow \mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} 1 \cdot \underbrace{\frac{1}{i}}_{\text{why?}} = O(\ln n). \square$$

Randomized algorithm for the hiring an student problem

To fool the input given by the adversary: Permute the input

```
Rand-Hire-Student (n)
Randomly permute the list [n]
best:=0
for i = 1 to n do
  interview i
  if i is better than best then
    best:=i and pre-select i
  end if
end for
```

Let $X(n)$ a r.v. counting the number of pre-selections, on an input of n students. Then

$$\mathbf{E}\left[X(n)\right] = O(\ln n)$$

Random-Quicksort

Consider the function Ran-Partition:

Ran-Partition $(A[p, \dots, q])$
$r = \text{rand}(p, q)$ u.a.r.
interchange $A[p]$ and $A[r]$

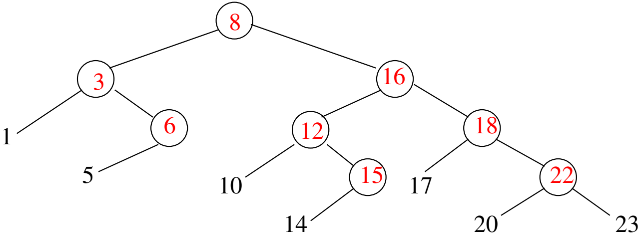
Using Ran-Partition, consider the following randomized D&C algorithm, on input $A[1, \dots, n]$:

Ran-Quicksort $(A[p, \dots, q])$
$r =$ **Ran-Partition** $(A[p, \dots, q])$
if $p < q$ **then**
  Ran-Quicksort $(A[1, \dots, r-1])$
  Ran-Quicksort $(A[r+1, \dots, q])$
else
  return $A[p]$
end if

Example

A={1,3,5,6,8,10,12,14,15,16,17,18,20,22,23}

Ran–Partition of input

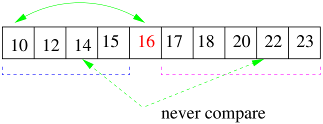


Expected Complexity of Ran-Partition

- Let X be a random variable counting the expected running time $T(n)$ of Rand-Quicksort. Notice the value of X is dominated by the number of comparisons.
- Every call to Rand-Partition has cost $\Theta(1) + \Theta(\underbrace{\text{number of comparisons}}_{p-q})$
- If we can count the number of comparisons, we can bound the the total time of Quicksort.
- Let X be the number of comparisons made in all calls of Ran-Quicksort
- X is a rv as it depends of the random choices of Ran-Partition

Expected Complexity of Ran-Partition

- Note: In the first application of Ran-Partition $A[r]$ compares with all $n-1$ elements.
- Key observation: Any two keys are compared iff one of them is a pivot, and they are compared at most one time.



For simplicity assume all keys are different, for any input $A[i,\ldots,j]$ to Ran-Quicksort, $1 \leq i < j \leq n$, let $Z_{i,j}$ be the ordered set of key $\{z_i, z_{i+1}, \ldots, z_j\}$ (with z_i the smallest).

- Note $|Z_{i,j}| = j - i + 1$
- Therefore choosing u.a.r. a pivot is done with probability

$$\frac{1}{|Z_{i,j}|} = \frac{1}{j-i+1}$$

.

Define the indicator r.v.:

$$X_{ij} = \begin{cases} 1 & \text{if } z_i \text{ is compared to } z_j, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{i,j}$
(this is true because we never compare a pair more than once)

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{i,j}\right] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbf{E}[X_{i,j}]$$

As $\mathbf{E}[X_{i,j}] = 0\mathbf{Pr}[X_{i,j} = 0] + 1\mathbf{Pr}[X_{i,j} = 1]$

$\therefore\ \mathbf{E}[X_{i,j}] = \mathbf{Pr}[X_{i,j} = 1] = \mathbf{Pr}[z_i \text{ is compared to } z_j]$

End of the proof and main theorem

$\mathbf{E}[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbf{E}[X_{i,j}] \, \mathbf{Pr}[z_i \text{ is compared to } z_j]$

As z_i and z_j compare iff one of them is chosen as pivot, then

$\mathbf{Pr}[X_{i,j}] = 1 = \mathbf{Pr}[z_i \text{ is pivot}] + \mathbf{Pr}[z_j \text{ is pivot}]$

Because pivots as chosen u.a.r. in $Z_{i,j}$:

$\mathbf{Pr}[z_i \text{ is pivot}] = \mathbf{Pr}[z_j \text{ is pivot}] = \frac{1}{j-i+1}$

Therefore:

$$\mathbf{E}[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1}.$$

$$
\begin{aligned}
\mathbf{E}[X] &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} \\
&= 2 \cdot \sum_{i=1}^{n} \left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-i+1}\right) \\
&< 2 \cdot \sum_{i=1}^{n} \left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) \\
&= 2 \cdot \sum_{i=1}^{n} H_n = 2 \cdot n \cdot H_n = 2n \ln n + \Theta(n).
\end{aligned}
$$

Theorem

The expected complexity of Ran-Quicksort is

$$\mathbf{E}[T(n)] = 2n \ln n + \Theta(n).$$

Selection and order statistics

Problem: Given a list A of n of unordered distinct keys, and a $i \in \mathbb{Z}, 1 \leq i \leq n$, select the element $x \in A$ that is larger than exactly $i-1$ other elements in A.

Notice if:

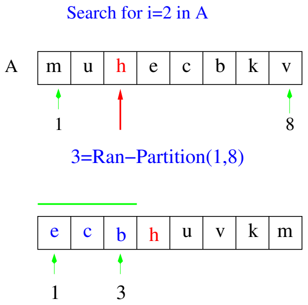
1. $i = 1 \Rightarrow$ MINIMUM element
2. $i = n \Rightarrow$ MAXIMUM element
3. $i = \lfloor \frac{n+1}{2} \rfloor \Rightarrow$ the MEDIAN
4. $i = \lfloor 0.9 \cdot n \rfloor \Rightarrow$ *order statistics*

Sort A ($O(n \lg n)$) and search for $A[i]$ ($\Theta(n)$).
Can we do it in linear time?
Yes, there is deterministic linear time algorithm for selection with a very bad constant.

Quick-Select

Given unordered $A[1, \ldots, n]$ return the i-th. element

- Quick-Select $(A[p, \ldots, q], i)$
- $r =$ Ran-Partition (p, q) to find position of pivot
- if $i = r$ return $A[r]$
- if $i < r$ Quick-Select $(A[p, \ldots, r-1], i)$
- else Quick-Select $(A[r+1, \ldots, q], i)$



Quick-Select Algorithm

Quickselect $(A[p, \ldots, q], i)$
if $p = q$ **then**
  return $A[p]$
else
  $r =$**Ran-Partition** $(A[p, \ldots, q])$
  $k = r - p + 1$
  if $i = k$ **then**
    return $A[q]$
    if $i < k$ **then**
      return Quickselect $(A[p, \ldots, q-1], i)$
    else
      return Quickselect $(A[q+1, \ldots, r], i-k)$
    end if
  end if
end if

Analysis.

- Lucky: at each recusrsive call the search space is reduced in 9/10 of the size. Then $T(n) \leq T(9n/10) + \Theta(n) = \Theta(n)$.
- Unlucky: $T(n) = T(n-1) + \Theta(n) = \Theta(n^2)$. In this case it is worst than sorting!.

Theorem

Given $A[1, \ldots, n]$ and i, the expected number of steps for Quick-Select to find the i-th. element in A is $O(n)$