CMPT 210: Probability and Computing

Lecture 16

Sharan Vaswani

March 12, 2024

Recap

Expectation/mean of a random variable R is denoted by $\mathbb{E}[R]$ and "summarizes" its distribution. Formally, $\mathbb{E}[R] := \sum_{\omega \in \mathcal{S}} \Pr[\omega]\, R[\omega]$

Alternate definition of expectation: $\mathbb{E}[R] = \sum_{x \in \mathrm{Range}(R)} x \Pr[R = x]$.

Linearity of Expectation: For n random variables $R_1, R_2, \ldots, R_n$ and constants $a_1, a_2, \ldots, a_n$, $\mathbb{E}\left[\sum_{i=1}^{n} a_i R_i\right] = \sum_{i=1}^{n} a_i\, \mathbb{E}[R_i]$.

Expectation - Examples

For a random variable $X : \mathcal{S} \to V$ and a function $g : V \to \mathbb{R}$, we define $\mathbb{E}[g(X)]$ as follows:

$$\mathbb{E}[g(X)] := \sum_{x \in \text{Range}(X)} g(x) \Pr[X = x]$$

If $g(x) = x$ for all $x \in \text{Range}(X)$, then $\mathbb{E}[g(X)] = \mathbb{E}[X]$.

Expectation - Examples

For a random variable $X : \mathcal{S} \to V$ and a function $g : V \to \mathbb{R}$, we define $\mathbb{E}[g(X)]$ as follows:

$$\mathbb{E}[g(X)] := \sum_{x \in \mathsf{Range}(X)} g(x) \Pr[X = x]$$

If $g(x) = x$ for all $x \in \mathsf{Range}(X)$, then $\mathbb{E}[g(X)] = \mathbb{E}[X]$.

Q: For a standard dice, if X is the r.v. corresponding to the number that comes up on the dice, compute $\mathbb{E}[X^2]$ and $(\mathbb{E}[X])^2$

Expectation - Examples

For a random variable $X : \mathcal{S} \to V$ and a function $g : V \to \mathbb{R}$, we define $\mathbb{E}[g(X)]$ as follows:

$$\mathbb{E}[g(X)] := \sum_{x \in \text{Range}(X)} g(x) \Pr[X = x]$$

If $g(x) = x$ for all $x \in \text{Range}(X)$, then $\mathbb{E}[g(X)] = \mathbb{E}[X]$.

Q: For a standard dice, if X is the r.v. corresponding to the number that comes up on the dice, compute $\mathbb{E}[X^2]$ and $(\mathbb{E}[X])^2$

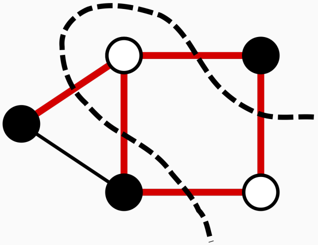
For a standard dice, $X \sim \text{Uniform}(\{1, 2, 3, 4, 5, 6\})$ and hence,

$$\mathbb{E}[X^2] = \sum_{x \in \{1,2,3,4,5,6\}} x^2 \Pr[X = x] = \frac{1}{6} \left[1^2 + 2^2 + \ldots + 6^2\right] = \frac{91}{6}$$

$$(\mathbb{E}[X])^2 = \left(\sum_{x \in \{1,2,3,4,5,6\}} x \Pr[X = x] \right)^2 = \left(\frac{1}{6} \left[1 + 2 + \ldots + 6\right] \right)^2 = \frac{49}{4}$$

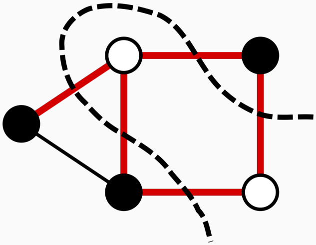
Max Cut

Given a graph $G = (\mathcal{V}, \mathcal{E})$, partition the graph's vertices into two complementary sets $\mathcal{S}$ and $\mathcal{T}$, such that the number of edges between the set $\mathcal{S}$ and the set $\mathcal{T}$ is as large as possible.



Max Cut

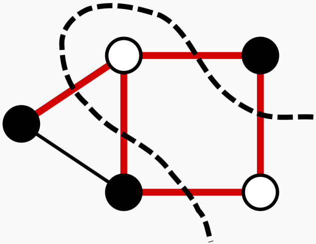
Given a graph $G = (\mathcal{V}, \mathcal{E})$, partition the graph's vertices into two complementary sets $\mathcal{S}$ and $\mathcal{T}$, such that the number of edges between the set $\mathcal{S}$ and the set $\mathcal{T}$ is as large as possible.



Max Cut has applications to VLSI circuit design.

Max Cut

Given a graph $G = (\mathcal{V}, \mathcal{E})$, partition the graph's vertices into two complementary sets $\mathcal{S}$ and $\mathcal{T}$, such that the number of edges between the set $\mathcal{S}$ and the set $\mathcal{T}$ is as large as possible.



Max Cut has applications to VLSI circuit design.

Equivalently, find a set $\mathcal{U} \subseteq \mathcal{V}$ of vertices that solve the following

$$\max_{\mathcal{U} \subseteq \mathcal{V}} |\delta(\mathcal{U})| \text{ where } \delta(\mathcal{U}) := \{(u, v) \in \mathcal{E} | u \in \mathcal{U} \text{ and } v \notin \mathcal{U}\}$$

Here, $\delta(\mathcal{U})$ is referred to as the "cut" corresponding to the set $\mathcal{U}$.

Max Cut

- Max Cut is NP-hard (Karp, 1972), meaning that there is no polynomial (in $|\mathcal{E}|$) time algorithm that solves Max Cut exactly.
- We want to find an approximate solution $\mathcal{U}$ such that, if OPT is the size of the optimal cut, then, $|\delta(\mathcal{U})| \geq \alpha\, OPT$ where $\alpha \in (0, 1)$ is the multiplicative approximation factor.
- Randomized algorithm that guarantees an approximate solution with $\alpha = \frac{1}{2}$ with probability close to 1 (Erdos, 1967).
- Algorithm with $\alpha = 0.878$. (Goemans and Williamson, 1995).
- Under some technical conditions, no efficient algorithm has $\alpha > 0.878$ (Khot et al, 2004).

Max Cut

- Max Cut is NP-hard (Karp, 1972), meaning that there is no polynomial (in $|\mathcal{E}|$) time algorithm that solves Max Cut exactly.
- We want to find an approximate solution $\mathcal{U}$ such that, if OPT is the size of the optimal cut, then, $|\delta(\mathcal{U})| \geq \alpha\, OPT$ where $\alpha \in (0,1)$ is the multiplicative approximation factor.
- Randomized algorithm that guarantees an approximate solution with $\alpha = \frac{1}{2}$ with probability close to 1 (Erdos, 1967).
- Algorithm with $\alpha = 0.878$. (Goemans and Williamson, 1995).
- Under some technical conditions, no efficient algorithm has $\alpha > 0.878$ (Khot et al, 2004).

We will use Erdos' randomized algorithm and first prove the result in expectation. We wish to prove that for $\mathcal{U}$ returned by Erdos' algorithm,

$$\mathbb{E}[|\delta(\mathcal{U})|] \geq \frac{1}{2} OPT$$

Max Cut

- Max Cut is NP-hard (Karp, 1972), meaning that there is no polynomial (in $|\mathcal{E}|$) time algorithm that solves Max Cut exactly.
- We want to find an approximate solution $\mathcal{U}$ such that, if OPT is the size of the optimal cut, then, $|\delta(\mathcal{U})| \geq \alpha\, OPT$ where $\alpha \in (0, 1)$ is the multiplicative approximation factor.
- Randomized algorithm that guarantees an approximate solution with $\alpha = \frac{1}{2}$ with probability close to 1 (Erdos, 1967).
- Algorithm with $\alpha = 0.878$. (Goemans and Williamson, 1995).
- Under some technical conditions, no efficient algorithm has $\alpha > 0.878$ (Khot et al, 2004).

We will use Erdos' randomized algorithm and first prove the result in expectation. We wish to prove that for $\mathcal{U}$ returned by Erdos' algorithm,

$$\mathbb{E}[|\delta(\mathcal{U})|] \geq \frac{1}{2} OPT$$

Algorithm: Select $\mathcal{U}$ to be a random subset of $\mathcal{V}$ i.e. for each vertex v, choose v to be in the set $\mathcal{U}$ independently with probability $\frac{1}{2}$ (do not even look at the edges!).

Max Cut

Claim: For Erdos' algorithm, $\mathbb{E}[|\delta(\mathcal{U})|] \geq \frac{1}{2} OPT$.

Max Cut

Claim: For Erdos' algorithm, $\mathbb{E}[|\delta(\mathcal{U})|] \geq \frac{1}{2} OPT$.

Proof: For each edge $(u, v) \in \mathcal{E}$, let $X_{u,v}$ be the indicator random variable equal to 1 iff the event $E_{u,v} = \{(u, v) \in \delta(\mathcal{U})\}$ happens.

Max Cut

Claim: For Erdos' algorithm, $\mathbb{E}[|\delta(\mathcal{U})|] \geq \frac{1}{2} OPT$.

Proof: For each edge $(u, v) \in \mathcal{E}$, let $X_{u,v}$ be the indicator random variable equal to 1 iff the event $E_{u,v} = \{(u, v) \in \delta(\mathcal{U})\}$ happens.

$$\mathbb{E}[|\delta(\mathcal{U})|] = \mathbb{E}\left[\sum_{(u,v)\in\mathcal{E}} X_{u,v}\right] = \sum_{(u,v)\in\mathcal{E}} \mathbb{E}\left[X_{u,v}\right] = \sum_{(u,v)\in\mathcal{E}} \Pr[E_{u,v}]$$

(Linearity of expectation, and Expectation of indicator r.v's.)

Max Cut

Claim: For Erdos' algorithm, $\mathbb{E}[|\delta(\mathcal{U})|] \geq \frac{1}{2} OPT$.

Proof: For each edge $(u, v) \in \mathcal{E}$, let $X_{u,v}$ be the indicator random variable equal to 1 iff the event $E_{u,v} = \{(u, v) \in \delta(\mathcal{U})\}$ happens.

$$\mathbb{E}[|\delta(\mathcal{U})|] = \mathbb{E}\left[\sum_{(u,v)\in\mathcal{E}} X_{u,v}\right] = \sum_{(u,v)\in\mathcal{E}} \mathbb{E}\left[X_{u,v}\right] = \sum_{(u,v)\in\mathcal{E}} \Pr[E_{u,v}]$$

(Linearity of expectation, and Expectation of indicator r.v's.)

$$\Pr[E_{u,v}] = \Pr[(u, v) \in \delta(\mathcal{U})] = \Pr\left[(u \in \mathcal{U} \cap v \notin \mathcal{U}) \cup (u \notin \mathcal{U} \cap v \in \mathcal{U})\right]$$

Max Cut

Claim: For Erdos' algorithm, $\mathbb{E}[|\delta(\mathcal{U})|] \geq \frac{1}{2} OPT$.

Proof: For each edge $(u, v) \in \mathcal{E}$, let $X_{u,v}$ be the indicator random variable equal to 1 iff the event $E_{u,v} = \{(u, v) \in \delta(\mathcal{U})\}$ happens.

$$\mathbb{E}[|\delta(\mathcal{U})|] = \mathbb{E}\left[\sum_{(u,v)\in\mathcal{E}} X_{u,v}\right] = \sum_{(u,v)\in\mathcal{E}} \mathbb{E}[X_{u,v}] = \sum_{(u,v)\in\mathcal{E}} \Pr[E_{u,v}]$$

(Linearity of expectation, and Expectation of indicator r.v's.)

$$\Pr[E_{u,v}] = \Pr[(u, v) \in \delta(\mathcal{U})] = \Pr[(u \in \mathcal{U} \cap v \notin \mathcal{U}) \cup (u \notin \mathcal{U} \cap v \in \mathcal{U})]$$

$$= \Pr[(u \in \mathcal{U} \cap v \notin \mathcal{U})] + \Pr[(u \notin \mathcal{U} \cap v \in \mathcal{U})] \quad \text{(Union rule for mutually exclusive events)}$$

Max Cut

Claim: For Erdos' algorithm, $\mathbb{E}[|\delta(\mathcal{U})|] \geq \frac{1}{2} OPT$.

Proof: For each edge $(u, v) \in \mathcal{E}$, let $X_{u,v}$ be the indicator random variable equal to 1 iff the event $E_{u,v} = \{(u, v) \in \delta(\mathcal{U})\}$ happens.

$$\mathbb{E}[|\delta(\mathcal{U})|] = \mathbb{E}\left[\sum_{(u,v)\in\mathcal{E}} X_{u,v}\right] = \sum_{(u,v)\in\mathcal{E}} \mathbb{E}\left[X_{u,v}\right] = \sum_{(u,v)\in\mathcal{E}} \Pr[E_{u,v}]$$

(Linearity of expectation, and Expectation of indicator r.v's.)

$$\Pr[E_{u,v}] = \Pr[(u, v) \in \delta(\mathcal{U})] = \Pr\left[(u \in \mathcal{U} \cap v \notin \mathcal{U}) \cup (u \notin \mathcal{U} \cap v \in \mathcal{U})\right]$$

$$= \Pr\left[(u \in \mathcal{U} \cap v \notin \mathcal{U})\right] + \Pr\left[(u \notin \mathcal{U} \cap v \in \mathcal{U})\right] \quad \text{(Union rule for mutually exclusive events)}$$

$$\Pr[E_{u,v}] = \Pr[u \in \mathcal{U}]\ \Pr[v \notin \mathcal{U}] + \Pr[u \notin \mathcal{U}]\ \Pr[v \in \mathcal{U}] = \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{1}{2}.$$

(Independent events)

Max Cut

Claim: For Erdos' algorithm, $\mathbb{E}[|\delta(\mathcal{U})|] \geq \frac{1}{2} OPT$.

Proof: For each edge $(u, v) \in \mathcal{E}$, let $X_{u,v}$ be the indicator random variable equal to 1 iff the event $E_{u,v} = \{(u, v) \in \delta(\mathcal{U})\}$ happens.

$$\mathbb{E}[|\delta(\mathcal{U})|] = \mathbb{E}\left[\sum_{(u,v)\in\mathcal{E}} X_{u,v}\right] = \sum_{(u,v)\in\mathcal{E}} \mathbb{E}\left[X_{u,v}\right] = \sum_{(u,v)\in\mathcal{E}} \Pr[E_{u,v}]$$

(Linearity of expectation, and Expectation of indicator r.v's.)

$$\Pr[E_{u,v}] = \Pr[(u, v) \in \delta(\mathcal{U})] = \Pr\left[(u \in \mathcal{U} \cap v \notin \mathcal{U}) \cup (u \notin \mathcal{U} \cap v \in \mathcal{U})\right]$$

$$= \Pr\left[(u \in \mathcal{U} \cap v \notin \mathcal{U})\right] + \Pr\left[(u \notin \mathcal{U} \cap v \in \mathcal{U})\right]$$ (Union rule for mutually exclusive events)

$$\Pr[E_{u,v}] = \Pr[u \in \mathcal{U}]\ \Pr[v \notin \mathcal{U}] + \Pr[u \notin \mathcal{U}]\ \Pr[v \in \mathcal{U}] = \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{1}{2}.$$

(Independent events)

$$\implies \mathbb{E}[|\delta(\mathcal{U})|] = \sum_{(u,v)\in\mathcal{E}} \Pr[E_{u,v}] = \frac{|\mathcal{E}|}{2} \geq \frac{\mathrm{OPT}}{2}.$$

Max Cut

Claim: For Erdos' algorithm, $\mathbb{E}[|\delta(\mathcal{U})|] \geq \frac{1}{2} OPT$.

Proof: For each edge $(u, v) \in \mathcal{E}$, let $X_{u,v}$ be the indicator random variable equal to 1 iff the event $E_{u,v} = \{(u, v) \in \delta(\mathcal{U})\}$ happens.

$$\mathbb{E}[|\delta(\mathcal{U})|] = \mathbb{E}\left[\sum_{(u,v)\in\mathcal{E}} X_{u,v}\right] = \sum_{(u,v)\in\mathcal{E}} \mathbb{E}\left[X_{u,v}\right] = \sum_{(u,v)\in\mathcal{E}} \Pr[E_{u,v}]$$

(Linearity of expectation, and Expectation of indicator r.v's.)

$$\Pr[E_{u,v}] = \Pr[(u, v) \in \delta(\mathcal{U})] = \Pr\left[(u \in \mathcal{U} \cap v \notin \mathcal{U}) \cup (u \notin \mathcal{U} \cap v \in \mathcal{U})\right]$$

$$= \Pr\left[(u \in \mathcal{U} \cap v \notin \mathcal{U})\right] + \Pr\left[(u \notin \mathcal{U} \cap v \in \mathcal{U})\right] \quad \text{(Union rule for mutually exclusive events)}$$

$$\Pr[E_{u,v}] = \Pr[u \in \mathcal{U}] \Pr[v \notin \mathcal{U}] + \Pr[u \notin \mathcal{U}] \Pr[v \in \mathcal{U}] = \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{1}{2}.$$

(Independent events)

$$\implies \mathbb{E}[|\delta(\mathcal{U})|] = \sum_{(u,v)\in\mathcal{E}} \Pr[E_{u,v}] = \frac{|\mathcal{E}|}{2} \geq \frac{\text{OPT}}{2}.$$

Conditional Expectation

Similar to probabilities, expectations can be conditioned on some event.

For random variable R, the expected value of R conditioned on an event A is given by:

$$\mathbb{E}[R|A] = \sum_{x \in \text{Range}(R)} x \Pr[R = x|A]$$

Conditional Expectation

Similar to probabilities, expectations can be conditioned on some event.

For random variable R, the expected value of R conditioned on an event A is given by:

$$\mathbb{E}[R|A] = \sum_{x \in \text{Range}(R)} x \Pr[R = x|A]$$

Q: If we throw a standard dice and define R to be the random variable equal to the number that comes up, what is the expected value of R given that the number is at most 4?

Conditional Expectation

Similar to probabilities, expectations can be conditioned on some event.

For random variable R, the expected value of R conditioned on an event A is given by:

$$\mathbb{E}[R|A] = \sum_{x \in \text{Range}(R)} x \Pr[R = x|A]$$

Q: If we throw a standard dice and define R to be the random variable equal to the number that comes up, what is the expected value of R given that the number is at most 4?

Let A be the event that the number is at most 4.
$\Pr[R = 1|A] = \frac{\Pr[(R=1) \cap A]}{\Pr[A]} = \frac{\Pr[R=1]}{\Pr[A]} = \frac{1/6}{4/6} = 1/4.$
$\Pr[R = 2|A] = \Pr[R = 3|A] = \Pr[R = 4|A] = \frac{1}{4}$ and $\Pr[R = 5|A] = \Pr[R = 6|A] = 0.$

$$\mathbb{E}[R|A] = \sum_{x \in \{1,2,3,4\}} x \Pr[R = x|A] = \frac{1}{4}[1 + 2 + 3 + 4] = \frac{5}{2}.$$

Conditional Expectation

Similar to probabilities, expectations can be conditioned on some event.

For random variable R, the expected value of R conditioned on an event A is given by:

$$\mathbb{E}[R|A] = \sum_{x \in \text{Range}(R)} x \Pr[R = x|A]$$

Q: If we throw a standard dice and define R to be the random variable equal to the number that comes up, what is the expected value of R given that the number is at most 4?

Let A be the event that the number is at most 4.
$\Pr[R = 1|A] = \frac{\Pr[(R=1) \cap A]}{\Pr[A]} = \frac{\Pr[R=1]}{\Pr[A]} = \frac{1/6}{4/6} = 1/4$.
$\Pr[R = 2|A] = \Pr[R = 3|A] = \Pr[R = 4|A] = \frac{1}{4}$ and $\Pr[R = 5|A] = \Pr[R = 6|A] = 0$.

$$\mathbb{E}[R|A] = \sum_{x \in \{1,2,3,4\}} x \Pr[R = x|A] = \frac{1}{4}[1 + 2 + 3 + 4] = \frac{5}{2}.$$

Q: What is the expected value of R given that the number is at least 4?

Law of Total Expectation

If R is a random variable $\mathcal{S} \to V$ and events $A_1, A_2, \ldots A_n$ form a partition of the sample space i.e. for all i, j, $A_i \cap A_j = \emptyset$ and $A_1 \cup A_2 \cup \ldots \cup A_n = \mathcal{S}$, then,

$$\mathbb{E}[R] = \sum_i \mathbb{E}[R|A_i] \Pr[A_i] \,.$$

Law of Total Expectation

If R is a random variable $\mathcal{S} \to V$ and events $A_1, A_2, \ldots A_n$ form a partition of the sample space i.e. for all i, j, $A_i \cap A_j = \emptyset$ and $A_1 \cup A_2 \cup \ldots \cup A_n = \mathcal{S}$, then,

$$\mathbb{E}[R] = \sum_i \mathbb{E}[R|A_i] \Pr[A_i] .$$

Proof:

$$\mathbb{E}[R] = \sum_{x \in \text{Range}(R)} x \Pr[R = x] = \sum_{x \in \text{Range}(R)} x \sum_i \Pr[R = x|A_i] \Pr[A_i]$$

(Law of total probability)

Law of Total Expectation

If R is a random variable $\mathcal{S} \to V$ and events $A_1, A_2, \ldots A_n$ form a partition of the sample space i.e. for all i, j, $A_i \cap A_j = \emptyset$ and $A_1 \cup A_2 \cup \ldots \cup A_n = \mathcal{S}$, then,

$$\mathbb{E}[R] = \sum_i \mathbb{E}[R|A_i] \Pr[A_i] \, .$$

Proof:

$$\mathbb{E}[R] = \sum_{x \in \text{Range}(R)} x \Pr[R = x] = \sum_{x \in \text{Range}(R)} x \sum_i \Pr[R = x|A_i] \Pr[A_i]$$

(Law of total probability)

$$= \sum_i \Pr[A_i] \sum_{x \in \text{Range}(R)} x \Pr[R = x|A_i]$$

Law of Total Expectation

If R is a random variable $\mathcal{S} \to V$ and events $A_1, A_2, \ldots A_n$ form a partition of the sample space i.e. for all i, j, $A_i \cap A_j = \emptyset$ and $A_1 \cup A_2 \cup \ldots \cup A_n = \mathcal{S}$, then,

$$\mathbb{E}[R] = \sum_i \mathbb{E}[R|A_i] \Pr[A_i] \,.$$

Proof :

$$\mathbb{E}[R] = \sum_{x \in \text{Range}(R)} x \Pr[R = x] = \sum_{x \in \text{Range}(R)} x \sum_i \Pr[R = x|A_i] \Pr[A_i]$$

(Law of total probability)

$$= \sum_i \Pr[A_i] \sum_{x \in \text{Range}(R)} x \Pr[R = x|A_i]$$

$$\implies \mathbb{E}[R] = \sum_i \Pr[A_i] \, \mathbb{E}[R|A_i].$$

Law of Total Expectation

If R is a random variable $\mathcal{S} \to V$ and events $A_1, A_2, \ldots A_n$ form a partition of the sample space i.e. for all i, j, $A_i \cap A_j = \emptyset$ and $A_1 \cup A_2 \cup \ldots \cup A_n = \mathcal{S}$, then,

$$\mathbb{E}[R] = \sum_i \mathbb{E}[R|A_i] \Pr[A_i] \,.$$

Proof:

$$\mathbb{E}[R] = \sum_{x \in \text{Range}(R)} x \Pr[R = x] = \sum_{x \in \text{Range}(R)} x \sum_i \Pr[R = x|A_i] \Pr[A_i]$$

(Law of total probability)

$$= \sum_i \Pr[A_i] \sum_{x \in \text{Range}(R)} x \Pr[R = x|A_i]$$

$$\implies \mathbb{E}[R] = \sum_i \Pr[A_i] \, \mathbb{E}[R|A_i].$$

Conditional Expectation - Examples

Q: Suppose that 49.6% of the people in the world are male and the rest female. If the expected height of a randomly chosen male is 5 feet 11 inches, while the expected height of a randomly chosen female is 5 feet 5 inches, what is the expected height of a randomly chosen person?

Conditional Expectation - Examples

Q: Suppose that 49.6% of the people in the world are male and the rest female. If the expected height of a randomly chosen male is 5 feet 11 inches, while the expected height of a randomly chosen female is 5 feet 5 inches, what is the expected height of a randomly chosen person?

Define H to be the random variable equal to the height (in feet) of a randomly chosen person. Define M to be the event that the person is male and F the event that the person is female. We wish to compute $\mathbb{E}[H]$ and we know that $\mathbb{E}[H|M] = 5 + \frac{11}{12}$ and $\mathbb{E}[H|F] = 5 + \frac{5}{12}$.
$\Pr[M] = 0.496$ and $\Pr[F] = 1 - 0.496 = 0.504$.
Hence, $\mathbb{E}[H] = \mathbb{E}[H|M] \Pr[M] + \mathbb{E}[H|F] \Pr[F] = \frac{71}{12}(0.496) + \frac{65}{12}(0.504)$.