ucc
Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

CS4618 Artificial Intelligence I

Today: Introduction to Probability (cont.)

Thomas Jansen

October 24th

Plans for Today

Introduction to Probability

Remember

- probability space is sample space (countable set) and probability measure (assigning probabilities to events)
- consequence event and counter-event $(\mathrm{Prob}(A) + \mathrm{Prob}\left(\overline{A}\right) = 1)$
- consequence union bound $(\mathrm{Prob}(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \mathrm{Prob}(A_1) + \mathrm{Prob}(A_2) + \cdots + \mathrm{Prob}(A_n))$
- independence of events (for all subsets $\mathrm{Prob}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathrm{Prob}(A_1) \cdot \mathrm{Prob}(A_2) \cdots \mathrm{Prob}(A_n))$
- random variable maps elementary events to numbers $(X : \Omega \to \mathbb{R}, \forall x_i \in \mathbb{R} : \mathrm{Prob}(X = x_i) = \mathrm{Prob}(\{s \in \Omega : X(s) = x_i\}))$
- expected value gives average value, mean $(\mathrm{E}(X) = \sum_{s \in \Omega} X(s) \cdot \mathrm{Prob}(s))$
- Observation Expected value may not correspond to possible

Direct Consequences

Theorem

Let X, Y be random variables on $(\Omega, Prob)$, $a \in \mathbb{R}$.

- $E(X) = \sum_x x \cdot Prob(X = x)$
- $E(a \cdot X) = a \cdot E(X)$
- $E(X + Y) = E(X) + E(Y)$

Proof directly from definition □

Consider random variable $X : \Omega \to \{0, 1\}$

Observation $\mathrm{E}(X) = \mathrm{Prob}(X = 1)$

Proof $\mathrm{E}(X) = 0 \cdot \mathrm{Prob}(X = 0) + 1 \cdot \mathrm{Prob}(X = 1)$
$= \mathrm{Prob}(X = 1)$ □

Direct Consequences (cont.)

Theorem

Let X, Y be independent random variables on $(\Omega, Prob)$.
$E(X \cdot Y) = E(X) \cdot E(Y)$

Proof

$$\begin{aligned}
&\mathrm{E}(X \cdot Y) = \sum_v v \cdot \mathrm{Prob}(X \cdot Y = v) \\
&= \sum_x \sum_y (x \cdot y) \cdot \mathrm{Prob}((X = x) \wedge (Y = y)) \\
&= \sum_x \sum_y (x \cdot y) \cdot \mathrm{Prob}(X = x) \cdot \mathrm{Prob}(Y = y) \\
&= \left(\sum_x x \cdot \mathrm{Prob}(X = x) \right) \cdot \left(\sum_y y \cdot \mathrm{Prob}(Y = y) \right) \\
&= \mathrm{E}(X) \cdot \mathrm{E}(Y)
\end{aligned}$$

□

Deviations from Expectation

Theorem (Markov Inequality)

$X \geq 0$ *random variable,* $s > 0$
$Prob\,(X \geq s \cdot E(X)) \leq \frac{1}{s}$

Proof Define $Y_t := \begin{cases} 1 & \text{if } X \geq t \\ 0 & \text{otherwise} \end{cases}$

Observe $X \geq t \cdot Y_t$

$$\Rightarrow \quad \mathrm{E}(X) \geq \mathrm{E}(t \cdot Y_t) = t \cdot \mathrm{E}(Y_t) = t \cdot \mathrm{Prob}(Y_t = 1)$$
$$= t \cdot \mathrm{Prob}(X \geq t)$$

$$\Leftrightarrow \quad \mathrm{Prob}(X \geq t) \leq \frac{\mathrm{E}(X)}{t}$$

Set $t := s \cdot \mathrm{E}(X)$ $\Rightarrow \mathrm{Prob}(X \geq s \cdot \mathrm{E}(X)) \leq \frac{1}{s}$ □

Binomial Distribution

Consider n independent fair coin tosses ($\Omega = \{\mathsf{H}, \mathsf{T}\}^n$)

Consider $M :=$ number of coins landing heads

What is $\mathrm{Prob}(M = i)$ and $\mathrm{E}(M)$?

Observation $$\mathrm{Prob}(M = i) = \begin{cases} \binom{n}{i} \cdot 2^{-n} & \text{if } i \in \{0, 1, \ldots, n\} \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow \quad \mathrm{E}(M) = \sum_{i=0}^{n} i \cdot \binom{n}{i} \cdot 2^{-n} = ?$$

Define $X_1, X_2, \ldots, X_n$ with $X_i := \begin{cases} 1 & \text{if } i\text{-th coin lands heads} \\ 0 & \text{otherwise} \end{cases}$

Clearly $\forall i \colon \mathrm{E}(X_i) = \frac{1}{2}$

Observation $$M = \sum_{i=1}^{n} X_i$$

$$\Rightarrow \quad \mathrm{E}(M) = \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \sum_{i=1}^{n} \mathrm{Prob}(X_i = 1) = \frac{n}{2}$$

Geometric Distribution

Consider Sequence of independent and identical unfair coin tosses for each coin $\text{Prob}(\text{coin lands heads}) = p$ with $p \in (0;1)$

Consider

$T :=$ number of coin tosses until the first coin lands heads

What is $\text{Prob}(T = i)$ and $\text{E}(T)$?

Observation $\forall i \in \mathbb{N}_0 : \text{Prob}(T = i) = p \cdot (1-p)^{i-1}$

$$\Rightarrow \quad \text{E}(T) = \sum_{i=0}^{\infty} i\text{Prob}(T = i) = \sum_{i=0}^{\infty} ip(1-p)^{i-1} = \sum_{i=1}^{\infty} ip(1-p)^{i-1}$$

$$= p\left(\sum_{i=1}^{\infty}\sum_{j=i}^{\infty}(1-p)^{j-1}\right) = p\left(\sum_{i=1}^{\infty}\left(\left(\sum_{j=1}^{\infty}(1-p)^{j-1}\right) - \left(\sum_{j=1}^{i-1}(1-p)^{j-1}\right)\right)\right)$$

$$= p\left(\sum_{i=1}^{\infty}\left(\left(\sum_{j=0}^{\infty}(1-p)^{j}\right) - \left(\sum_{j=0}^{i-2}(1-p)^{j}\right)\right)\right)$$

$$= p \cdot \left(\sum_{i=1}^{\infty}\frac{1}{1-(1-p)} - \frac{1-(1-p)^{i-1}}{1-(1-p)}\right) = p \cdot \left(\sum_{i=1}^{\infty}\frac{(1-p)^{i-1}}{p}\right)$$

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

Example Independent Fair Coin Tosses

Consider 100 independent fair coin tosses

Consider $M :=$ number of coins landing heads

We know $\mathrm{E}(M) = \frac{1}{2} \cdot 100 = 50$

What is $\mathrm{Prob}(M \geq 75)$?

Markov
$$\begin{aligned}\mathrm{Prob}(M \geq 75) &= \mathrm{Prob}\left(M \geq \tfrac{3}{2} \cdot 50\right) \\ &= \mathrm{Prob}\left(M \geq \tfrac{3}{2} \cdot \mathrm{E}(M)\right) \leq \tfrac{2}{3}\end{aligned}$$

Observation weak bound
$$\begin{aligned}\mathrm{Prob}(M \geq 75) &= \sum_{i=75}^{100} \mathrm{Prob}(M = i) = \sum_{i=75}^{100} \binom{100}{i} \cdot 2^{-100} \\ &= \frac{89{,}310{,}453{,}796{,}450{,}805{,}935{,}325}{316{,}912{,}650{,}057{,}057{,}350{,}374{,}175{,}801{,}344} < 0.000000282\end{aligned}$$

Deviations from Expectation Revisted

Theorem (Chernoff Bounds)

Let $X_1, X_2, \ldots, X_n : \Omega \to \{0,1\}$ *independent random variables with*

$\forall i \in \{1,2,\ldots,n\} : 0 < Prob(X_i = 1) < 1.$

Let $X := \sum_{i=1}^{n} X_i.$

$\forall \delta > 0 : Prob(X > (1+\delta) \cdot E(X)) < \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{E(X)}$

$\forall 0 < \delta < 1 : Prob(X < (1-\delta) \cdot E(X)) < e^{-E(X)\delta^2/2}$

Proof Ideas

- Transform X to $e^{t \cdot X}$ for some $t > 0$.
- Apply Markov Inequality.
- Exploit independence of $X_1, \ldots, X_n$.

Proof of Chernoff Bounds (1)

Let $t \in \mathbb{R}^+$.

Observe $\quad \text{Prob}\left(X > (1+\delta)\text{E}\left(X\right)\right) = \text{Prob}\left(e^{tX} > e^{t(1+\delta)\text{E}(X)}\right)$

Markov $\rightsquigarrow$ $\text{Prob}\left(X > (1+\delta)\text{E}\left(X\right)\right) < \frac{\text{E}\left(e^{tX}\right)}{e^{t(1+\delta)\text{E}(X)}}$

Remember $\quad X = \sum_{i=1}^{n} X_i$

Thus $\quad \text{E}\left(e^{tX}\right) = \text{E}\left(e^{t\cdot\sum_{i=1}^{n} X_i}\right) = \text{E}\left(\prod_{i=1}^{n} e^{tX_i}\right)$

Remember $\quad X_i$ completely independent

$\rightsquigarrow e^{tX_i}$ completely independent

Thus $\quad \text{E}\left(e^{tX}\right) = \prod_{i=1}^{n} \text{E}\left(e^{tX_i}\right)$

Observe $\quad \text{E}\left(e^{tX_i}\right) = \text{Prob}\left(X_i = 0\right)\cdot 1 + \text{Prob}\left(X_i = 1\right)\cdot e^t$

$= \text{Prob}\left(X_i = 1\right)\cdot e^t + 1 - \text{Prob}\left(X_i = 1\right)$

$= 1 + \text{E}\left(X_i\right)\cdot\left(e^t - 1\right) \leq e^{\text{E}(X_i)\cdot\left(e^t-1\right)}$

Proof of Chernoff Bounds (2)

We have $$\text{Prob}\left(X > (1+\delta)\text{E}\left(X\right)\right) < \frac{\prod_{i=1}^{n} e^{\text{E}(X_i)\cdot\left(e^t-1\right)}}{e^{t(1+\delta)\text{E}(X)}}$$

$$= \frac{e^{\left(e^t-1\right)\cdot\sum_{i=1}^{n}\text{E}(X_i)}}{e^{t(1+\delta)\text{E}(X)}} = \frac{e^{\left(e^t-1\right)\cdot\text{E}(X)}}{e^{t(1+\delta)\text{E}(X)}}$$

We have $\delta > 0$, thus $t := \ln(1+\delta)$ feasible

This yields $$\text{Prob}\left(X > (1+\delta)\text{E}\left(X\right)\right) < \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\text{E}(X)}$$ ✓

Analogously with $t := \ln\left(\frac{1}{1-\delta}\right)$

$$\text{Prob}\left(X < (1-\delta)\text{E}\left(X\right)\right) < \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\text{E}(X)}$$

Observation $0 < \delta < 1 \Rightarrow (1-\delta)^{1-\delta} > e^{-\delta+\delta^2/2}$

$\Rightarrow$ $$\text{Prob}\left(X < (1-\delta)\text{E}\left(X\right)\right) < e^{-\text{E}(X)\delta^2/2}$$ □

Example Independent Fair Coin Tosses Revisited

Consider 100 independent fair coin tosses
$M :=$ number of coins landing heads $(\mathrm{E}(M) = 50)$

What is $\mathrm{Prob}(M \geq 75)$?

We know Markov $\mathrm{Prob}(M \geq 75) \leq \frac{2}{3}$
truth $\mathrm{Prob}(M \geq 75) < 0.000000282$

Can we apply Chernoff bounds?

Observe random variables $X_1, X_2, \ldots, X_{100}$ with

$$X_i := \begin{cases} 1 & i\text{-th coin lands heads} \\ 0 & \text{otherwise} \end{cases}$$

yields $M = \sum_{i=1}^{n} X_i$, $X_1, X_2, \ldots, X_n$ completely independent

$$\mathrm{Prob}(M \geq 75) = \mathrm{Prob}\left(M \geq \left(1 + \tfrac{1}{2}\right) \cdot \mathrm{E}(M)\right) < \left(\frac{\sqrt{e}}{(3/2)^{(3/2)}}\right)^{50}$$
$$< 0.0044722$$

Taking Knowledge Into Account

Consider events $A_1, A_2, \ldots, A_n, B$

What changes if we know that B occured?

Clearly new probability distribution such that
$\forall s \in \Omega \setminus B \colon \tilde{P}(s) = 0$

Still need $\sum_{s \in \Omega \cap B} \tilde{P}(s) = 1$

Moreover relative probabilities inside B should not change
$\forall A_i, A_j \subseteq B \colon \frac{\text{Prob}(A_i)}{\text{Prob}(A_j)} \stackrel{!}{=} \frac{\tilde{P}(A_i)}{\tilde{P}(A_j)}$

Definition

For events A, B with $B \neq \emptyset$ define the conditional probability of A given B as $\text{Prob}(A \mid B) := \frac{\text{Prob}(A \cap B)}{\text{Prob}(B)}$.

Direct Consequences

Theorem (Law of Total Probability)

Let B_i with $i \in I$ be a partition of Ω.

$$\forall A \subseteq \Omega\colon Prob(A) = \sum_{i \in I} Prob(A \mid B_i) \cdot Prob(B_i)$$

Proof

$$\text{Prob}(A \mid B_i) \cdot \text{Prob}(B_i) = \frac{\text{Prob}(A \cap B_i)}{\text{Prob}(B_i)} \cdot \text{Prob}(B_i) = \text{Prob}(A \cap B_i)$$

$$\sum_{i \in I} \text{Prob}(A \mid B_i) \cdot \text{Prob}(B_i) = \sum_{i \in I} \text{Prob}(A \cap B_i)$$

$$= \text{Prob}\left(\bigcup_{i \in I} (A \cap B_i)\right) = \text{Prob}\left(A \cap \bigcup_{i \in I} B_i\right) = \text{Prob}(A) \qquad \square$$

Corollary

$$\forall A, B \subseteq \Omega\colon Prob(A) \geq Prob(B) \cdot Prob(A \mid B)$$

Law of Total Probability for Random Variables

Clearly for random variable X, event B, $\mathrm{E}(X \mid B)$ well defined

Theorem

Let B_i with $i \in I$ be a partition of Ω, X a random variable.
$E(X) = \sum_{i \in I} E(X \mid B_i) \cdot Prob(B_i)$

Proof Remember $\mathrm{E}(X) = \sum_{s \in \Omega} X(s) \cdot \mathrm{Prob}(s)$

$$\mathrm{E}(X \mid B_i) \cdot \mathrm{Prob}(B_i) = \sum_{s \in \Omega} X(s) \cdot \mathrm{Prob}(s \mid B_i) \cdot \mathrm{Prob}(B_i)$$
$$= \sum_{s \in \Omega} X(s) \cdot \mathrm{Prob}(s \cap B_i)$$

Observation $\forall s \in \Omega \colon \sum_{i \in I} \mathrm{Prob}(s \cap B_i) = \mathrm{Prob}(s)$

since B_i are a partition

$$\sum_{i \in I} \mathrm{E}(X \mid B_i) \cdot \mathrm{Prob}(B_i) = \sum_{i \in I} \sum_{s \in \Omega} X(s) \cdot \mathrm{Prob}(s \cap B_i)$$
$$= \sum_{s \in \Omega} \left(X(s) \sum_{i \in I} \mathrm{Prob}(s \cap B_i) \right) = \sum_{s \in \Omega} X(s) \mathrm{Prob}(s) = \mathrm{E}(X) \qquad \square$$

Summary & Take Home Message

Things to remember

- Markov inequality
- distributions: binomial, geometric
- Chernoff bounds
- conditional probability

Take Home Message

- Probability theory is not very complicated.
- Proving useful results is not very difficult.