Goal:- basic probability refresher.

For you to do:-
- Refresh your probability
- Read the additional notes
- Try out basic exercises.

Probability Review

ECE 4200

Machine Learning and Probability

- Machine learning:
  - Learning phase: gain information about the system given training data
  - Inference phase: using the information learned to infer about new samples (test data).
  - Often the process generating data is not deterministic:?

- examples :- data may not be deterministic .
  9999...
-

What is probability?

Imagine a process that has many outcomes. Every time you run the process one of the outcomes is possible.

The relative chance of each outcome is its probability.

$\{1, \dots 6\}$

Example:

If you roll a fair die, the probability of obtaining 4 is 1/6.

Real life: No fair die, estimate likelihoods

An Example

Example:

You throw a coin 1000 times, you see 314 heads. What is the likelihood that the next throw will be a head?

0.314. → $\boxed{\text{Maximum Likelihood principle.}}$

$\boxed{\text{H, H,}}$

H ←

Definitions

- Outcome space: Ω
  - Die: $\Omega = \{1,2,3,4,5,6\}$, Coin: $\Omega = \{H, T\}$
- Event: a subset of Ω $\{2,3,5\}$.
- $\mathcal{E}$: set of all possible events
  - If discrete, $\underbrace{|\mathcal{E}| = 2^{|\Omega|}}$? $\longrightarrow$ $\{1,\ldots,6\}$. $\longrightarrow$ 64 $\leftarrow 2^6$

A probability distribution p is a function from Ω to $[0,1]$, s.t.

$$\sum_{x \in \Omega} p(x) = 1.$$

When Ω is continuous, summation is replaced with integral.

Examples

- For a fair die,

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}.$$

$$|\mathcal{E}| = 64$$

- Example: E: outcome is a prime number

$$p(2) + p(3) + p(5) = \frac{1}{2}.$$

Continuous distributions

- What is the temperature at noon?
- Infinite possibilities:

$$|\Omega| = [0, \infty) \text{ (in Kelvin)}$$

- What is the probability that it will be more than 50F?

$$E = \{Temp > 50F\}?$$

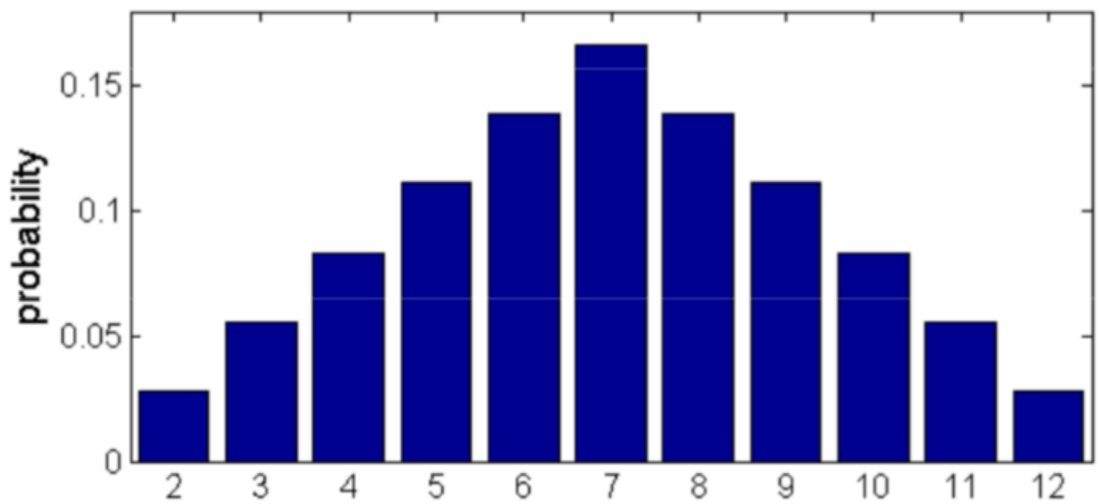
$$p(E) = \int_{x \in E} p(x) dx.$$

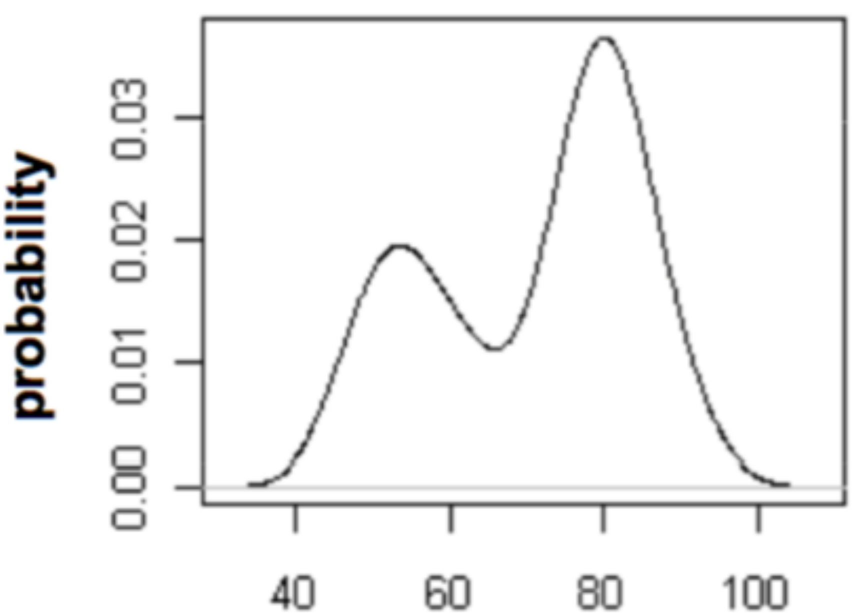
Continuous distributions

- Suppose p is a uniform distribution over $[0,2]$. What is the probability of the outcome being more than 1.5?

PMF, pdf

probability mass function (pmf)





Rules of probability

1. For any set $E \subseteq \Omega$,

$$p(E) = \sum_{x \in \mathrm{E}} p(x)\,.$$

2. $p(\Omega) = 1$ $\quad \sum_{x \in \Omega} p(x)$

3. For E_1, E_2, such that $E_1 \cap E_2 = \emptyset$,

$$p(E_1 \cup E_2) = p(E_1) + p(E_2).$$

4. $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$.

Marginalization

The probability

- For disjoint events $B_1, B_2 \ldots$, such that $B_1 \cup B_2 \ldots = \Omega$.

$$p(E) = \sum_i p(E \cap B_i)\,.$$

Prove this.

Conditional probability

The probability

- $p(E_1|E_2)$: Probability that E_1 happens, given E_2 happens.

$$p(E_1|E_2) = \frac{p(E_1 \cap E_2)}{p(E_2)}.$$

$$p(E_1 \cap E_2) = p(E_2)p(E_1|E_2).$$

Ex

What is the probability that the outcome of a roll is even, given it is a prime?

E_1: outcome is even, E_2: outcome is prime.

BB.

Independence

Two events E_1 and E_2 are independent if

$$p(E_1 \cap E_2) = p(E_1)p(E_2).$$

$$p(E_1|E_2) = \frac{p(E_1 \cap E_2)}{p(E_2)} = p(E_1).$$

$$p(E_2|E_1) = \frac{p(E_1 \cap E_2)}{p(E_1)} = p(E_2)..$$

Chain Rule

The probability

- For events, $E_1, E_2 \ldots, E_k$,

$$p(E_1 \cap \cdots \cap E_k) = p(E_1)p(E_2|E_1)p(E_3|\underbrace{E_1 \cap E_2}) \ldots$$

Can now work with many different events.

If $\boxed{E_1, E_2 \ldots, E_k}$ are mutually independent, then:

$$p(E_1 \cap \cdots \cap E_k) = p(E_1)p(E_2) \ldots p(E_k)$$

Bayes Rule

The probability

$$p(E_1|E_2) = \frac{p(E_1 \cap E_2)}{p(E_2)}.$$

Marginalization, and conditional probability:

$$p(E_2) = p(E_2 \cap E_1) + p(E_2 \cap E_1^c)$$

marginalization

$$= p(E_1)p(E_2|E_1) + p(E_1^c)p(E_2|E_1^c).$$

Bayes Rule:

$$p(E_1|E_2) = \frac{p(E_1)p(E_2|E_1)}{p(E_1)p(E_2|E_1) + p(E_1^c)p(E_2|E_1^c)}.$$

→ **Example :-**

Monty Hall Example

BB.

Expectation

X: Outcome of a random process, a random variable.

Expectation of a random variable:

$$\boxed{\mathbb{E}[X] = \sum_x x \cdot p(x)\,.}$$

$\int_x x \, p(x) dx$

$X_1, X_2, \ldots, X_k$: $\mathbb{E}[X_1 + \cdots X_k] = \mathbb{E}[X_1] + \cdots \mathbb{E}[X_k]$

Linearity of expectations. Proof?

Prove that expected roll of a fair die is 3.5.

Variance

How concentrated is a random variable?

Variance: How much deviation of a random variable around its mean?

$$\sum_x (x - E(x))^2 \cdot p(x)$$

$$\boxed{Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].}$$

Compute the variance of a roll of a fair die.

Show that : $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Entropy

Which one of the random variables is more “random”?

$$X_1: \quad p(X=0) = p(X=1) = 0.5$$

$$X_2: \quad p(X=0) = 0 \qquad p(X=1) = 1$$

$$X_3: \quad p(X=0) = 0.3 \qquad p(X=1) = 0.7$$

Entropy

What about these?

$$X_1: \quad p(X=0)=0.3, p(X=1)=0.2, p(X=2)=0.5$$

$$X_2: \quad p(X=0)=0.33, p(X=1)=0.25, p(X=\overset{2}{\cancel{1}})=0.4$$

$$X_3: \quad p(X=0)=0\,, \qquad p(X=1)=0, \qquad p(X=\overset{2}{\cancel{1}})=1$$

$$X_4: \quad p(X=0)=0.33\,, p(X=1)=0.33, p(X=\overset{2}{\cancel{1}})=0.33$$

Entropy

Ω - discrete

How to measure the randomness of a variable? $(0, \infty)$

$\lim_{p(x) \to 0} p(x) \log \frac{1}{p(x)} = 0$

$$H(X) := \sum_{x \in \Omega} p(x) \log_2 \frac{1}{p(x)}$$

$\boxed{0 \cdot \log \frac{1}{0} = 0}$

X_1: $p(X = 0) = p(X = 1) = 0.5$ $H(X_1) = 1$

X_2: $p(X = 0) = 0$ $p(X = 1) = 1$ $H(X_2) = 0$

X_3: $p(X = 0) = 0.25$ $p(X = 1) = 0.75$ $H(X_2) = 0.81$

Entropy

What about these?

$$X_1: \quad p(X = 0) = 0.3, p(X = 1) = 0.2, p(X = 2) = 0.5$$

$$X_2: \quad p(X = 0) = 0.35, p(X = 1) = 0.25, p(X = 2) = 0.4$$

$$X_3: \quad p(X = 0) = 0\,, \qquad p(X = 1) = 0, \qquad p(X = 2) = 1$$

$$X_4: \quad p(X = 0) = 0.33\,, p(X = 1) = 0.33, p(X = 2) = 0.33$$

$$H(X_1) = 1.49, H(X_2) = 1.56, H(X_3) = 0, H(X_4) = 1.59$$

Properties of Entropy

Suppose X is a discrete random variable over Ω and U is a uniform random variable over Ω.

$$0 \leq H(X) \leq \ H(U) = \log|\Omega|$$

$$H(X) = 0 \text{ if and only if } \exists\ x \in \Omega, p(x) = 1$$

Information Gain:

Let S be the set of n elements, and n_i is the number of elements with label i, then

$$H(S) = \frac{n_i}{n}\sum_{i} \log\frac{n}{n_i}$$

Let A be an attribute, and S_x is the set of elements in S with $A = x$. Then:

$$IG(S, A) = H(S) - \sum_{x \in A} \frac{|S_x|}{|S|} H(S_x)$$