Data 100, Discussion 6 – Probability and Modelling

Suraj Rampure

March 8, 2019

Random Variables

A **random variable** is a variable whose values are determined by probabilities. In other words, a random variable is a function that takes outcomes of a random process to real numbers.

Usually, we use X or Y to denote a random variable.

Probability Mass Functions

The **probability mass function**, or **density function**, tells us the specific probabilities of each event occurring. In general, these can be discrete or continuous, but for the purposes of this class they will be discrete.

$P(\overset{\text{r.v.}}{X} = \overset{\text{specific outcome}}{x})$ represents the probability of random variable X taking on the value of x. (x is normally numerical, but it doesn't necessarily have to be.) $\mathbb{X}$, then, represents the set of the actual values themselves (i.e. dice rolls, cards, etc.) $\mathcal{X} : \{1, 2, 3, 4, 5, 6\}$ $x \in \mathcal{X}$

A few properties that $P(X = x)$ must satisfy:

$$0 \leq P(X = x) \leq 1, \forall x \in \mathbb{X}$$

$$\sum_{x \in \mathbb{X}} P(X = x) = 1$$

Expectation

Often, we want to determine the expected, or average, outcome of a probabilistic event. For that, we define the **expectation** of a random variable:

$$\mathbb{E}[X] = \sum_{x \in \mathbb{X}} x \cdot P(X = x) \quad = \sum \text{outcome} \cdot P(\text{outcome})$$

sum over all possible outcomes

This is a weighted average of all possible outcomes.

Properties that the expectation satisfies:

$$E[c] = c \quad , \forall c \in \mathbb{R}$$

$$\mathbb{E}[aX] = a\mathbb{E}[X]$$

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y], \forall\, r.v.\ X, Y$$

Independent

$$P(X=x, Y=y) = P(X=x)\,P(Y=y)$$

$$P(A \cap B) = P(A) \cdot P(B)$$

independent $\Longrightarrow$ uncorrelated

but not vice versa!

X : fair, 6 sided die

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6}$$

$$= 3.5$$

Variance

We define variance as being the mean squared distance from the mean, i.e.

$$var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

$$E[(X-\mu)^2]$$

Properties that variance satisfies:

$$var[aX] = a^2 var[X]$$

$$\text{if X, Y are independent} : var[X+Y] = var[X] + var[Y]$$

$$E[g(X)] = \sum_{x \in X} g(x) \cdot P(X=x)$$

$$E[X^2] = \sum x^2 P(X=x)$$

Bernoulli Random Variables

A Bernoulli random variable is one that only assumes two possible values – 0 or 1. It has a single **parameter**, p.

$$P(X=1)=p$$

$$P(X=0)=1-p$$

$$\sum P(X=x) = P(X=0) + P(X=1) = 1-p+p = 1$$

$$\mathbb{E}[X]=0\cdot P(X=0)+1\cdot P(X=1)=p$$

$$var[X]=p(1-p)$$

Bernoulli Random Variables

A Bernoulli random variable is one that only assumes two possible values – 0 or 1. It has a single **parameter**, p.

$$P(X=1)=p$$

$$P(X=0)=1-p$$

$$\mathbb{E}[X]=0\cdot P(X=0)+1\cdot P(X=1)=p$$

$$var[X]=p(1-p)$$