Random variables (discrete)

Saad Mneimneh

1 Introducing random variables

A random variable is a mapping from the sample space to the real line. We usually denote the random variable by X, and a value that it can take by x.

Example 1: Rolling a die

The sample space is $\{., .., ..., ..., ..., ...\}$. We define a random variable by mapping the dots in the outcome to the number they represent. Our random variable can take the following values $\{1, 2, 3, 4, 5, 6\}$.

$$P(X = x) = 1/6 \text{ for } x \in \{1, 2, 3, 4, 5, 6\}$$

In other words,

$$P(X = 1) = P(X = 2) = ... = P(X = 6) = 1/6$$

Example 2: Tossing a coin

The sample space is $\{H, T\}$. We define a random variable by mapping T to 0 and H to 1. Our random variable can take the following values $\{0, 1\}$.

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

where p is the probability of getting a head (1/2 if coin is fair).

2 Probability mass function

A probability mass function (PMF) assigns a probability to each value of the random variable.

2.1 Uniform PMF

For instance, the following is a PMF for the die example above.

```
     |
     |
1/6--|----o----o----o----o----o----o--
     |    |    |    |    |    |    |
     |    |    |    |    |    |    |
   __|____|____|____|____|____|____|_____
          1    2    3    4    5    6
```

Such a PMF is called uniform (discrete uniform), because all probabilities are equal.

2.2 Binomial PMF

Consider tossing the coin n times. If $n = 3$, for instance, the sample space is the following (8 possible outcomes):

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

In general, we have 2^n possible outcomes. If the coin is fair, all outcomes are equally likely, with each having a probability of $1/2^n$ (uniform). If $P(H) = 1 - P(T) = p$, then outcomes have different probabilities. The probability of an outcome is given by $p^k(1-p)^{n-k}$ where k is the number of heads in the outcome.

Consider the random variable that maps each outcome to the number of heads. Then, $P(X = k)$ is $p^k(1-p)^{n-k}$ multiplied by the number of outcomes with k heads.

$$P(X = k) = b(k, n, p) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}$$

where

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n!}{k!(n-k)!}$$

is the number of ways of choosing k elements from a set of n elements. This is equivalent to the number of outcomes with k heads (we need to choose k tosses out of n tosses and make them heads, the remaining $n - k$ tosses are automatically tails). The above PMF is known as the binomial PMF.

2.3 Geometric PMF

Now, consider tossing the coin until getting a head. The sample space is (infinite) the following:

$$S = \{H, TH, TTH, TTTH, TTTTH, TTTTTH, TTTTTTH, TTTTTTTH, \ldots\}$$

For a given outcome,

$$P(T.......TH) = p(1-p)^k$$

where k is the number of tails in the outcome.

Define a random variable that maps each outcome to the number of tosses. Then, it is obvious that:

$$P(X = k) = p(1-p)^{k-1}$$

This is knows as the geometric PMF, because: $P(X = 1) = p(1-p)^0$, $P(X = 2) = p(1-p)^1$, $P(X = 3) = p(1-p)^2$, $P(X = 4) = p(1-p)^3$,... The probability decreases geometrically in $(1-p)$. Note that

$$\sum_k P(X = k) = p[1 + (1-p) + (1-p)^2 + (1-p)^3 + ...] = \frac{p}{1-(1-p)} = 1$$

3 Expectation (mean) of a random variable

The expected value (or mean) of a random variable X is defined as (like a weighted average):

$$E[X] = \sum_x xP(X = x)$$

For instance, the expected value of a uniform random variable defined over the set $\{1, 2, ..., n\}$ is:

$$E[X] = \sum_{i=1}^{n} xP(X = x) = 1/n \cdot \sum_{i=1}^{n} x$$

$$= 1/n \cdot (1 + 2 + 3 + ... + n) = 1/n \cdot n(n+1)/2 = (n+1)/2$$

Expectation can also be defined for any function $f(x)$, $f(x) = x$ above being a special case. In this case, we use the notation

$$E[f(x)] = \sum_x f(x)P(X = x)$$

Expectation has the following properties:

- The expected value of a constant is the constant itself.
- $E[aX] = aE[X]$, where a is a constant.
- Linearity (very useful): $E[X + Y] = E[X] + E[Y]$, regardless of whether the two random variables X and Y are independent or not.
- $E[XY] = E[X]E[Y]$ if X and Y are independent.

- Conditional expectation: $E[X] = E[E[X|Y = y]]$, the inner expectation is over X using $P(X|Y = y)$ which results in an expression in y, the outer is the expectation of that expression over Y using $P(Y)$.

Here's an example of using linearity of expectation. If X is a binomial random variable, then X can be viewed as the sum of independent random variables $X_1 + X_2 + \ldots + X_n$, where each X_i is the following random variable (recall the coin example):

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

As defined above, X is called a Bernoulli trial. Then $E[X] = 1 \cdot p + 0 \cdot (1 - p) = p$. Using linearity of expectation,

$$E[X] = E[X_1 + \ldots + X_n] = E[X_1] + ... + E[X_n] = nE[X] = np$$

The fact that these Bernoulli trials are independent is not relevant to the above computation.

Here's an example of using conditional expectation: Let X be a Bernoulli trial with parameter p as above; however, p is unknown. Assume that p is coming from a uniform distribution such that $P(p = 0.25) = P(p = 0.5) = P(p = 0.75) = 1/3$. Then

$$E[X] = E[E[X|p]] = E[np] = nE[p] = n/2$$

If X is a geometric random variable, then $E[X] = 1/p$ (this can be shown by explicit evaluation of the sum as given in the formula for expectation). If, however, p is unknown and is coming from the same distribution above, then $E[X] = E[E[X|p]] = E[1/p] = 1/0.25 \cdot 1/3 + 1/0.5 \cdot 1/3 + 1/0.75 \cdot 1/3 = 22/9$. Note that $E[1/p]$ is not the same as $1/E[p]$.

4 Variance of a random variable

The variance of a random variable X is defined as:

$$\sigma_X^2 = E[(X - E[X])^2] \text{ (positive deviation from the mean)}$$

The quantity $E[(X - c)^2]$ is minimized when $c = E[X]$ (this is easy to show). Therefore, the variance is that minimum. One can also show that, given two random variables X and Y, $E[(X - c)^2|Y = y]$ is minimized when $c = E[X|Y = y]$.

Using properties of expectation, we can rewrite the above as:

$$\sigma_X^2 = E[X^2] - E[X]^2$$

We can compute the following:

- Uniform: $\sigma_X^2 = (n^2 - 1)/12$
- Binomial: $\sigma_X^2 = np(1 - p)$
- Geometric: $\sigma_X^2 = (1 - p)/p^2$

It is easy to verify that the variance of a constant is zero, and that $\sigma_{aX}^2 = a^2\sigma_X^2$. Consider two random variables X and Y and let us compute the variance of their sum.

$$\sigma_{X+Y}^2 = E[(X + Y)^2] - E[(X + Y)]^2$$
$$= E[X^2] + E[Y^2] + 2E[XY] - E[X]^2 - E[Y]^2 - 2E[X]E[Y]$$
$$= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 + 2(E[XY] - E[X]E[Y])$$
$$= \sigma_X^2 + \sigma_Y^2 + 2(E[XY] - E[X]E[Y])$$

Note that if X and Y are indepdendent then $E[XY] = E[X]E[Y]$; therefore, $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$. The term $Cov(X, Y) = E[XY] - E[X]E[Y] = E[(X - E[X])(Y - E[Y])]$ is called the covariance of X and Y. If the covariance is zero, we say that the random variables are uncorrelated (not necessarily independent though). Here's an example of two random variables X and Y that are uncorrelated but dependent. Let $P(X = i) = P(X = -i) = p_i$, for $i = 0, 1, 2, \ldots$. Let $Y = X^2$. It is easy to show that $E[X] = E[X^3] = 0$. Therefore, $E[XY] - E[X]E[Y] = E[X^3] - E[X]E[Y] = 0 - 0 \cdot E[Y] = 0$. However, X and Y are not independent.