Week 10 Random Variable and Expectation

Tuesday, 26 May 2020 2:53 PM

Random Variables

Definition

An (integer) **random variable** is a function from Ω to $\mathbb{Z}$.
In other words, it associates a number value with every outcome.

Random variables are often denoted by $X, Y, Z, \ldots$

Example

Random variable $X_s \stackrel{\text{def}}{=}$ sum of rolling two dice

$\Omega = \{(1,1), (1,2), \ldots, (6,6)\}$

$X_s((1,1)) = 2 \qquad X_s((1,2)) = 3 = X_s((2,1)) \quad \ldots$

Expectation

Definition

The **expected value** (often called "expectation" or "average") of a random variable X is

$$E(X) = \sum_{k \in \mathbb{Z}} P(X = k) \cdot k$$

Theorem (linearity of expected value)

$E(X + Y) = E(X) + E(Y)$
$E(c \cdot X) = c \cdot E(X)$

Standard Deviation and Variance

Definition

For random variable X with expected value (or: **mean**) $\mu = E(X)$, the **standard deviation** of X is

$$\sigma = \sqrt{E((X-\mu)^2)}$$

and the **variance** of X is

$$\sigma^2$$

Standard deviation and variance measure how spread out the values of a random variable are. The smaller σ^2 the more confident we can be that $X(\omega)$ is close to $E(X)$, for a randomly selected ω.

NB

The variance can be calculated as $E((X-\mu)^2) = E(X^2) - \mu^2$

Cumulative Distribution Functions

Definition

The **cumulative distribution function** $\mathrm{CDF}_X : \mathbb{Z} \longrightarrow \mathbb{R}$ of an integer random variable X is defined as

$$\mathrm{CDF}_X(y) \mapsto \sum_{k \leq y} P(X = k)$$

$\mathrm{CDF}_X(y)$ collects the probabilities $P(X)$ for all values up to y

Example: Binomial Distributions

Definition

Binomial random variables count the number of 'successes' in n independent experiments with probability p for each experiment.

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

$$\mathrm{CDF}_B(y) \mapsto \sum_{k \leq y} \binom{n}{k} p^k (1-p)^{n-k}$$

Theorem

If X is a binomially distributed random variable based on n and p, then $E(X) = n \cdot p$ with variance $\sigma^2 = n \cdot p \cdot (1-p)$

Normal Distribution

Fact

For large n, binomial distributions can be approximated by **normal distributions** *(a.k.a.* **Gaussian distributions***) with mean $\mu = n \cdot p$ and variance $\sigma^2 = n \cdot p \cdot (1-p)$*

$$\frac{1}{\sqrt{2\sigma^2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The expected sum when rolling two dice is

$\boxed{9.3.3}$ Buy one lottery ticket for \$1. The only prize is \$1M. Each ticket has probability $6 \cdot 10^{-7}$ of winning.

The expected sum when rolling two dice can be computed as

Example

$E(S_n)$, where $S_n \stackrel{\text{def}}{=} |\text{no. of HEADS in } n \text{ tosses}|$

- 'hard way'

$$E(S_n) = \sum_{k=0}^{n} P(S_n = k) \cdot k = \sum_{k=0}^{n} \frac{1}{2^n} \binom{n}{k} \cdot k$$

since there are $\binom{n}{k}$ sequences of n tosses with k HEADS, and each sequence has the probability $\frac{1}{2^n}$

$$= \frac{1}{2^n} \sum_{k=1}^{n} \frac{n}{k} \binom{n-1}{k-1} k = \frac{n}{2^n} \sum_{k=0}^{n-1} \binom{n-1}{k} = \frac{n}{2^n} \cdot 2^{n-1} = \frac{n}{2}$$

using the 'binomial identity' $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

- 'easy way'

$$E(S_n) = E(S_1^1 + \ldots + S_1^n) = \sum_{i=1\ldots n} E(S_1^i) = nE(S_1) = n \cdot \frac{1}{2}$$

Note: $S_n \stackrel{\text{def}}{=} |\text{HEADS in } n \text{ tosses}|$ while each $S_1^i \stackrel{\text{def}}{=} |\text{HEADS in 1 toss}|$

You face a quiz consisting of six true/false questions, and your plan is to guess the answer to each question (randomly, with probability 0.5 of being right). There are no negative marks, and answering four or more questions correctly suffices to pass.
What is the probability of passing and what is the expected score?

$\boxed{9.3.7}$

An urn has $m + n = 10$ marbles, $m \geq 0$ red and $n \geq 0$ blue.
7 marbles selected at random without replacement.
What is the expected number of red marbles drawn?

Find the average waiting time for the first HEAD, with no upper bound on the 'duration' (one allows for all possible sequences of tosses, regardless of how many times TAILS occur initially).

$$A = E(X_w) = \sum_{k=1}^{\infty} k \cdot P(X_w = k) = \sum_{k=1}^{\infty} k \frac{1}{2^k}$$
$$= \frac{1}{2^1} + \frac{2}{2^2} + \frac{3}{2^3} + \dots$$

$\boxed{9.4.12}$ A die is rolled until the first 4 appears. What is the expected waiting time?

$P(\text{roll } 4) = \frac{1}{6}$ hence $E(\text{no. of rolls until first } 4) = 6$

Example

To find an object $\mathcal{X}$ in an unsorted list L of elements, one needs to search linearly through L. Let the probability of $\mathcal{X} \in L$ be p, hence there is $1 - p$ likelihood of $\mathcal{X}$ being absent altogether. Find the expected number of comparison operations.

But what about his performance?

- If one of his numbers comes up, say a_i, he wins \$35 from the bet on that number and loses \$23 from the bets on the remaining numbers, thus collecting \$12.
  This happens with probability $p = \frac{24}{37}$.
- With probability $q = \frac{13}{37}$ none of his numbers appears, leading to loss of \$24.

The expected result

$\boxed{\text{9.5.10 (supp)}}$ Two independent experiments are performed.
$P(\text{1st experiment succeeds}) = 0.7$
$P(\text{2nd experiment succeeds}) = 0.2$
Random variable X counts the number of successful experiments.

(a) Expected value of X? $E(X) = 0.7 + 0.2 = 0.9$

(b) Probability of exactly one success? $0.7 \cdot 0.8 + 0.3 \cdot 0.2 = 0.62$

(c) Probability of at most one success? $(\text{b}) + 0.3 \cdot 0.8 = 0.86$

(e) Variance of X? $\sigma^2 = (0.62 \cdot 1 + 0.14 \cdot 4) - 0.9^2 = 0.37$

$\boxed{9.4.10}$ An experiment is repeated 30,000 times with probability of success $\frac{1}{4}$ each time.

(a) Expected number of successes? $E(X) = 30,000 \cdot \frac{1}{4} = 7500$

(b) Standard deviation? $\sigma = \sqrt{30,000 \cdot \frac{1}{4} \cdot \frac{3}{4}} = 75$