Random Variables

•In some experiments, we would like to assign a numerical value to each possible outcome in order to facilitate a mathematical analysis of the experiment.

•For this purpose, we introduce **random variables**.

•**Definition:** A random variable is a function from the sample space of an experiment to the set of real numbers. That is, a random variable assigns a real number to each possible outcome.

•Note: Random variables are functions, not variables, and they are not random, but map random results from experiments onto real numbers in a well-defined manner.

Random Variables

•**Example:**

•Let X be the result of a rock-paper-scissors game.
•If player A chooses symbol a and player B chooses symbol b, then

•X(a, b) = 1, if player A wins,
• = 0, if A and B choose the same symbol,
• = -1, if player B wins.

Random Variables

•A(rock, rock) =	•0
•A(rock, paper) =	•-1
•A(rock, scissors) =	•1
•A(paper, rock) =	•1
•A(paper, paper) =	•0
•A(paper, scissors) =	•-1
•A(scissors, rock) =	•-1
•A(scissors, paper) =	•1
•A(scissors, scissors) =	•0

Expected Values

•Once we have defined a random variable for our experiment, we can statistically analyze the outcomes of the experiment.

•For example, we can ask: What is the **average value** (called the **expected value**) of a random variable when the experiment is carried out a large number of times?

•Can we just calculate the arithmetic mean across all possible values of the random variable?

Expected Values

•No, we cannot, since it is possible that some outcomes are more likely than others.

•For example, assume the possible outcomes of an experiment are 1 and 2 with probabilities of 0.1 and 0.9, respectively.

•Is the average value 1.5?

•No, since 2 is much more likely to occur than 1, the average must be larger than 1.5.

Expected Values

•Instead, we have to calculate the **weighted sum** of all possible outcomes, that is, each value of the random variable has to be multiplied with its probability before being added to the sum.

•In our example, the average value is given by $0.1 \cdot 1 + 0.9 \cdot 2 = 0.1 + 1.8 = 1.9$.

•**Definition:** The **expected value** (or expectation) of the random variable X(s) on the sample space S is equal to:

•$E(x) = \sum_{s \in S} p(s)X(s)$.

Expected Values

•**Example:** Let X be the random variable equal to the sum of the numbers that appear when a pair of dice is rolled.

•There are **36 outcomes** (= pairs of numbers from 1 to 6).

•The **range** of X is {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}.

•Are the 36 outcomes equally likely?

•Yes, if the dice are not biased.

•Are the 11 values of X equally likely to occur?

•No, the probabilities vary across values.

Expected Values

•P(X = 2) = 1/36
•P(X = 3) = 2/36 = 1/18
•P(X = 4) = 3/36 = 1/12
•P(X = 5) = 4/36 = 1/9
•P(X = 6) = 5/36
•P(X = 7) = 6/36 = 1/6
•P(X = 8) = 5/36
•P(X = 9) = 4/36 = 1/9
•P(X = 10) = 3/36 = 1/12
•P(X = 11) = 2/36 = 1/18
•P(X = 12) = 1/36

Expected Values

•$E(X) = 2\cdot(1/36) + 3\cdot(1/18) + 4\cdot(1/12) + 5\cdot(1/9) + 6\cdot(5/36) + 7\cdot(1/6) + 8\cdot(5/36) + 9\cdot(1/9) + 10\cdot(1/12) + 11\cdot(1/18) + 12\cdot(1/36)$

•$E(X) = 7$

•This means that if we roll the dice many times, sum all the numbers that appear and divide the sum by the number of trials, we expect to find a value of 7.

Expected Values

•**Theorem 3, p. 480 (p. 429 6th ed.):**

if X_i, i = 1, 2, …, n with a positive integer n, are random variables on S, then

$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$.

•Moreover, if a and b are real numbers, then

$E(aX + b) = aE(X) + b$.

(proof of Theorem 3)

- $E(X+Y) = \sum_{s\in S} p(s)\{X+Y\}(s)$ (def. of E(X))
  $= \sum_{s\in S} p(s)(X(s)+Y(s))$
  $= \sum_{s\in S} p(s)X(s) + \sum_{s\in S} p(s)Y(s)$
  $= E(X) + E(Y)$
- $E(aX + b) = \sum_{s\in S} p(s)(aX(s)+b)$
  $= \sum_{s\in S} p(s)aX(s) + \sum_{s\in S} p(s)b$
  $= a\sum_{s\in S} p(s)X(s) + b\sum_{s\in S} p(s)$
  $= aE(X) + b$. (a and b real numbers)

Expected Values

•Knowing this theorem, we could now solve the previous example much more easily:

•Let X_1 and X_2 be the numbers appearing on the first and the second die, respectively.

•For each die, there is an equal probability for each of the six numbers to appear. Therefore, $E(X_1) = E(X_2) = (1 + 2 + 3 + 4 + 5 + 6)/6 = 7/2$.

•We now know that
$E(X_1 + X_2) = E(X_1) + E(X_2) = 7$.

Expected Values

- We can use our knowledge about expected values to compute the average-case complexity of an algorithm.
- Let the sample space be the set of all possible inputs $a_1, a_2, \ldots, a_n$, and the random variable X assign to each a_j the number of operations that the algorithm executes for that input.
- For each input a_j, the probability that this input occurs is given by $p(a_j)$.
- The algorithm's average-case complexity then is:
- $E(X) = \sum_{j=1,\ldots,n} p(a_j)X(a_j)$

Expected Values

- However, in order to conduct such an average-case analysis, you would need to find out:
  - the number of steps that the algorithms takes for any (!) possible input, and
  - the probability for each of these inputs to occur.
- For most algorithms, this would be a highly complex task, so we will stick with the worst-case analysis.
- On page 483 in the textbook (page 481 in the 6th Edition), an average-case analysis of the linear search algorithm is shown.

Binomial Distribution

- A random variable X is **Binomial (n,p)** if $P(X = j) = C(n,j)p^j q^{n-j}$, for j = 0,1,2,…,n. Here $0 \le p \le 1$, and q = 1-p.
- Note that by the Binomial Theorem $\Sigma_{j=0}^{n} P(X=j) = (p+q)^n = 1$.
- We can give two proofs that E(X) = np. See Theorem 2, page 479 (p 428, 6th ed) for the direct proof. (easier proof coming up….)

Binomial Distribution

- Note that X is Binomial (n,p) if X counts the number of successes in n independent trials of an event with probability p of success and q of failure on any trial.

Expectation of a Binomial R.V.

- If X is binomial(n,p), let X_i be 1 if the i^{th} trial gives a success, 0 otherwise.
- Then $E(X_i) = 1p + 0q = p$ and $X = \Sigma_{i=1}^{n} X_i$ (X is the number of successes in the n trials) so $E(X) = \Sigma_{i=1}^{n} E(X_i) = np$.
- For a direct proof see Theorem 2, p 479

Expectation

- We defined $E(X) = \sum_{s \in S} p(s)X(s)$.
- But also, $E(X) = \sum_{r \in X(S)} r\, P(X = r)$.
- In the first sum, we sum over all outcomes s the value of X at s weighted by the prob. of s.
- In the second sum we sum over all values X takes on, grouping all outcomes s such that X(s) = r in the event X=r.
- P(X=r) is the prob. of that set of outcomes, and r is X(s) for each s in the event X=r.

An Expectation example

- Roll a die. S = {1,2,3,4,5,6} (the outcomes)
- Define the random variable X by X(1) = X(2) = X(3) = 1, X(4) = X(5) = 6, X(6) = 18.
- $E(x) = \sum_{s\in S} p(s)X(s) = \sum_{s=1}^{6} (1/6)X(s)$
- = (1/6)*1 + (1/6)*1 + (1/6)*1 + (1/6)*6 + (1/6)*6 + (1/6)*18
- = (3/6)*1+ (2/6)*6 + (1/6)*18 = 5.5
- Note: the last sum is just $\sum_{r\in X(S)} r\, P(X = r)$.

Independent Random Variables

•**Definition:** The random variables X and Y on a sample space S are **independent** if

•$p(X(s) = u \wedge Y(s) = v) = p(X(s) = u) \cdot p(Y(s) = v)$.

•In other words, X and Y are independent if the probability that $X(s) = u \wedge Y(s) = v$ equals the product of the probability that X(s) = u and the probability that Y(s) = v for all real numbers u and v.

•This means that the events "X(s) = u" and "Y(s) = v" are independent for every u and v.

Independent Random Variables

•**Example:** Are the random variables X_1 and X_2 from the "pair of dice" example independent?

•**Solution:**

•$p(X_1 = i) = 1/6$

•$p(X_2 = j) = 1/6$

•$p(X_1 = i \wedge X_2 = j) = 1/36$

•Since $p(X_1 = i \wedge X_2 = j) = p(X_1 = i)\cdot p(X_2 = j)$, the random variables X_1 and X_2 are **independent**.

Independent Random Variables

•**Theorem:** If X and Y are independent random variables on a sample space S, then E(XY) = E(X)E(Y).

•**Note:**

•E(X + Y) = E(X) + E(Y) is true for any X and Y, but

•E(XY) = E(X)E(Y) needs X and Y to be independent.

•How come?

- Proof: the proof is subtle.
- $E(XY) = \Sigma_{s\in S} X(s)Y(s)p(s)$, sum over outcomes.
- $= \Sigma_{u\in X(S), v\in Y(S)}\, uvP(X{=}u \text{ and } Y{=}v)$, sum over values X & Y take on, grouping outcomes.
- $= \Sigma_{u\in X(S), v\in Y(S)}\, uvP(X{=}u)P(Y{=}v)$, since X,Y are independent.
- $= (\Sigma_{u\in X(S)} uP(X{=}u))(\Sigma_{v\in Y(S)} vP(Y{=}v))$
- =E(X)E(Y)

Independent Random Variables

•**Example:** Let X and Y be random variables on the sample space, and each of them assumes the values 1 and 3 with equal probability.

•Then E(X) = E(Y) = 2

•If X and Y are **independent**, we have:

•$E(X + Y) = 1/4\cdot(1 + 1) + 1/4\cdot(1 + 3) + 1/4\cdot(3 + 1) + 1/4\cdot(3 + 3) = 4 = E(X) + E(Y)$

•$E(XY) = 1/4\cdot(1\cdot1) + 1/4\cdot(1\cdot3) + 1/4\cdot(3\cdot1) + 1/4\cdot(3\cdot3) = 4 = E(X)\cdot E(Y)$

Independent Random Variables

- Let us now assume that X and Y are **correlated** in such a way that Y = 1 whenever X = 1, and Y = 3 whenever X = 3.
- $E(X + Y) = 1/2\cdot(1 + 1) + 1/2\cdot(3 + 3)$
  $= 4 = E(X) + E(Y) = 2 + 2$
- $E(XY) = 1/2\cdot(1\cdot1) + 1/2\cdot(3\cdot3)$
  $= 5 \neq E(X)\cdot E(Y) = 2\cdot2$
- So, we can guarantee the average value of XY to be the average value of X * the average value of Y **<u>if</u>** X and Y are independent

Variance

- The **expected value** of a random variable is an important parameter for the description of a random distribution.
- It does not tell us, however, anything about **how widely distributed** the values are.
- This is described, at least in part, by the **variance** of a random variable.

Variance

- **Definition:** Let X be a random variable on a sample space S. The **variance** of X, denoted by V(X), is
- $V(X) = \sum_{s\in S}(X(s) - E(X))^2 p(s)$.
- The **standard deviation** of X, denoted by $\sigma(X)$, is defined to be the square root of V(X).
- A large variance means the distribution is spread out, a small variance means it is more localized.

Variance

- **Useful rules:**
- If X is a random variable on a sample space S, then $V(X) = E(X^2) - E(X)^2$. $V(aX) = a^2V(X)$
- If X and Y are two independent random variables on a sample space S, then V(X + Y) = V(X) + V(Y).
- Furthermore, if X_i, i = 1, 2, …, n, with a positive integer n, are pairwise independent random variables on S, then
  $V(X_1 + X_2 + \ldots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n)$.
- Proofs coming up, and in the textbook on page 489 (6th edition 436, 437).

Variance

- Theorem. If X is a random variable on a sample space S, then
  (a) $V(X) = E(X^2) - E(X)^2$
  (b) $V(aX) = a^2V(X)$
- Proof: (b)
  $V(aX) = \sum_{s\in S}(aX(s) - E(aX))^2 p(s)$
  $= \sum_{s\in S}(aX(s) - aE(X))^2 p(s) = a^2V(X)$

- Proof (a): $V(X) = \sum_{s\in S}(X(s) - E(X))^2 p(s)$
  $= \sum_{s\in S}(X(s)^2 - 2E(X)X(s) + E(X)^2)p(s)$
  $= \sum_{s\in S}X(s)^2p(s) - 2E(X)\sum_{s\in S}X(s)p(s) + E(X)^2\sum_{s\in S}p(s)$
  $= E(X^2) - 2E(X)E(X) + E(X)^2$
  $= E(X^2) - E(X)^2$

Variance

- Theorem (p 489)
  If X and Y are independent random variables on a sample space S, then
  $V(X + Y) = V(X) + V(Y)$.
- Generalizing, for independent rv's $X_1,\ldots,X_n$, $V(\sum_{j=1}^{n} X_j) = \sum_{j=i}^{n} V(X_j)$

- Proof: recall that if X and Y are independent random variables then $E(XY) = E(X)E(Y)$.
- Thus, $V(X+Y) = E((X+Y)^2) – E(X+Y)^2$
- $= E(X^2 + 2XY + Y^2) – (E(X) + E(Y))^2$
- $= E(X^2) + 2E(X)E(Y) + E(Y^2) +$
  $-E(X)^2 - 2E(X)E(Y) – E(y)^2$
- $= E(X^2) - E(X)^2 + E(Y^2) - E(y)^2$
- $= V(X) + V(Y)$

Variance

- **So:**
- If X is a random variable on a sample space S, $V(X) = E(X^2) – E(X)^2$. $V(aX) = a^2V(X)$
- If X and Y are two independent random variables on a sample space S, then $V(X + Y) = V(X) + V(Y)$.
- Thus, if X and Y are independent and have the same variance, then $V(X+Y) = 2V(X)$
- Now, if X = Y, then X and Y are far from independent, and $V(X+Y) = V(2X) = 4V(X)$

Binomial Distribution

- Theorem: If X is binomial (n,p) then
  $E(X) = np$
  $V(X) = npq$
- Proof: We proved $E(X) = np$ earlier.
- If $X_i = 1$ if the i^{th} trial is a success and 0 otherwise then $E(X_i) = p$, independent RVs.
- $V(X_i) = E(X_i^2) – p^2 = p – p^2 = p(1-p) = pq$.
- But $X = \Sigma_{i=1}^{n} X_i$, so $V(X) = \Sigma_{i=1}^{n} V(X_i) = npq$

Geometric Distribution

- Def. A r.v. X has the geometric distribution with parameter p if $P(X=k) = (1-p)^{k-1}p$, k = 1,2,3,4,…
- Example: X could be the number of times you have to flip a coin before getting an H, if P(H) = p on any flip.
- Note: the geometric distribution has infinitely many values, but is discrete.
- Theorem. If X is geometric with parameter p, then
  $E(X) = 1/p$, $V(X) = (1-p)/p^2$

Geometric Distribution

- let $f(x) = \Sigma_{n=0}^{\infty} x^n = (1-x)^{-1}$. Then:
- 1. $f'(x) = \Sigma_{n=1}^{\infty} nx^{n-1} = (1-x)^{-2}$, and
- 2. $f''(x) = \Sigma_{n=2}^{\infty} n(n-1)x^{n-2} = 2(1-x)^{-3}$.
- $E(X) = \Sigma_{n=1}^{\infty} nP(X=n) = \Sigma_{n=1}^{\infty} n(1-p)^{n-1}p$
  $= p(1-(1-p))^{-2} = 1/p$, using 1.
- $V(X) = \Sigma_{n=1}^{\infty} (n-p^{-1})^2 P(X=n)$
  $= \Sigma_{n=1}^{\infty} (n-p^{-1})^2 (1-p)^{n-1}p$
  $= \Sigma_{n=1}^{\infty} (n^2-2np^{-1}+p^{-2})(1-p)^{n-1}p$
  $= \Sigma_{n=1}^{\infty} (n(n-1) +n -2np^{-1}+p^{-2})(1-p)^{n-1}p$
  $= \Sigma_{n=1}^{\infty} (n(n-1) +n(1-2p^{-1})+p^{-2})(1-p)^{n-1}p$
  $= (1-p)p\Sigma_{n=2}^{\infty} n(n-1)(1-p)^{n-2} + (1-2p^{-1})p\Sigma_{n=1}^{\infty} n (1-p)^{n-1} +$
  $p^{-2} \Sigma_{n=1}^{\infty} (1-p)^{n-1}p$
  $=(1-p)p2p^{-3} + (p-2)p^{-2} + p^{-2}$, using 2, 1, & sum of all probs.
  $= p^{-2}(1-p)$