Probability

June 27, 2004

1 Events and probability

Example

Suppose we roll a fair six-sided die. The set of possible outcomes is $\Omega = \{1, 2, 3, 4, 5, 6\}$. We can consider many possible events, e.g. "the result is 5", "the result is at least 4", "the result is divisible by 2".

Since the die is fair we would say that all 6 outcomes are equally likely, with each having probability 1/6. So writing P for Probability, $P(\text{result is } i) = 1/6$ for $i = 1, 2, \dots, 6$ and $P(\text{result is at least 4}) = P(4, 5, 6) = 3/6$.

1.1 Outcomes and events

Consider an experiment with a set of possible outcomes Ω.

- The set of outcomes is called the *sample space*.
- A particular outcome $\omega \in \Omega$ is called a *sample point*.
- An *event* is a subset of Ω. An event A is said to occur if the outcome ω satisfies $\omega \in A$.

Examples

1. Tossing a coin. $\Omega = \{H, T\}$. Event "getting a head": $A = \{H\}$.
2. Rolling a die twice. $\Omega = \{(i, j) : i, j = 1, 2, \dots, 6\}$. Event "obtaining a total of 4": $A = \{(1, 3), (2, 2), (3, 1)\}$.
3. Measuring the lifetime of a lightbulb. $\Omega = [0, \infty)$. Possible event "the bulb is still working after t time units": $A = (t, \infty)$.
4. Record the price of a share over a period of trading of length T. $\Omega = \{f : [0, T] \to [0, \infty]\}$. Possible event "the price never falls below L": $A = \{f : [0, T] \to [L, \infty)\}$.

For example 2, if the outcome of the experiment is $(2, 2)$ then then event A occurs.

For events $A, B \subseteq \Omega$,

- $A \cap B$ corresponds to the event "A and B".
- $A \cup B$ corresponds to the event "A or B or both".

- A^C ($\equiv \Omega \backslash A$) corresponds to the event "A does not occur".
- $A \backslash B$ ($\equiv A - B$) corresponds to the event "A but not B".

1.2 Elementary probability

If the sample space Ω is finite and if each sample point $\omega \in \Omega$ is equally likely then we can consider the special case in which the probability of an event A is defined to be $P(A) = \frac{|A|}{|\Omega|}$.

Example

A random number generator generates R random digits. For $k = 0, 1, \ldots, 9$ find the probability that:

1. no digit exceeds k.
2. k is the greatest digit generated.

Take the first sentence to mean that $\Omega = \{(d_1, d_2, \ldots, d_r) : d_i = 0, 1, \ldots, 9; i = 1, 2, \ldots, r\}$ with each of the 10^r sample points equally likely.

1. The event of interest $A_k = \{(d_1, d_2, \ldots, d_r) : d_i = 0, 1, \ldots, k; i = 1, 2, \ldots, r\}$. Here $|A_k| = (k+1)^r$, so $P(A_k) = \frac{|A_k|}{|\Omega|} = \frac{(k+1)^r}{10^r}$.
2. Write B_k for the event that k is the greatest digit generated. $B_k = A_k \backslash A_{k-1}$ (with $A_{-1} = \emptyset$). Now $A_{k-1} \subseteq A_k$ so $|B_k| = |A_k| - |A_{k-1}| = (k+1)^r - k^r$. Thus $P(B_k) = \frac{|B_k|}{|\Omega|} = \frac{(k+1)^r - k^r}{10^r}$.

1.3 Counting

Ordered selection

Suppose we have n balls numbered $1, \ldots, n$ in a box and that we choose them sequentially. There are $n!$ possible outcomes. If only r of the balls are chosen, the number of possible outcomes is $n(n-1)\ldots(n-r+1)$. This procedure is called *sampling without replacement*.

If the balls are retuned to the box before the next choice is made, the procedure is called *sampling with replacement*[1]. The number of possible outcomes when r choices are made is n^r.

Examples

In a group of r people, what's the probability that two or more have the same birthday? Write b_i for the birthday (day of year) of the ith person. Then $\Omega = \{(b_1, \ldots, b_r) : b_i = 1, 2, \ldots, 365; i = 1, 2, \ldots, r\}$. Assume that all 365^r outcomes are equally likely. Write $A = \{\text{two or more people share a birthday}\}$. Then $A^C = \{\text{all } r \text{ people have different birthdays}\} = 365 \times 364 \times \ldots \times (365 - r + 1) = 365^{\underline{r}}$. Since $|A| = |\Omega| - |A^C| = 365^r - 365^{\underline{r}}$, $P(A) = 1 - \frac{365^{\underline{r}}}{365^r}$.

[1] surprisingly enough

Unordered selection

Recall that no. of subsets of $\{1, \ldots, n\}$ with r elements is $\binom{n}{r} = \frac{n!}{r!(n-r)!}$. More generally there are $\frac{n!}{n_1!n_2!\ldots n_m!}$ ways of partitioning the set $\{1, \ldots, n\}$ into a first n_1-subset, a second n_2-subset, $\ldots$, and an mth n_m-subset, where $\sum_{k=1}^{m} n_k = n$.

Example

What is the probability of the event A that a hand in bridge (13 cards) contains 5 spades, 4 hearts, 3 diamonds and 1 club? No. of hands of cards $= \binom{52}{13}$. No. of hands with 5S, 4H, 3D, 1C is $\binom{13}{5}\binom{13}{4}\binom{13}{3}\binom{13}{1}$, so

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{13}{5}\binom{13}{4}\binom{13}{3}\binom{13}{1}}{\binom{52}{13}}.$$

1.4 Probability measures

A collection $\mathcal{F}$ of subsets of Ω is called an *event space* or a σ-field if

1. $\Omega, \emptyset \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$
3. $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Note that since $A \cap B = (A^C \cup B^C)^C$, $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$ and more generally $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

The informal idea is that in a particular application with sample space Ω, the event space $\mathcal{F}$ corresponds to the collection of events "of interest". (We will se that "of interest" means those sets whose probabilities we may wish to know or calculate.)

Examples

1. $\mathcal{F} = \{\text{all subsets of } \Omega\}$
2. $\mathcal{F} = \{\emptyset, \Omega\}$ or $\mathcal{F} = \{\emptyset, A, A^C, \Omega\}$ for some $A \in \Omega$.

Definition A function $P : \mathcal{F} \to \mathbf{R}$ is called a probability measure if

I $0 \le P(A) \le 1 \quad \forall A \in \mathcal{F}$

II $P(\Omega) = 1$

III If $A_1, A_2, \ldots$ are disjoint events in $\mathcal{F}$ then $P(\bigcup_{i=1}^{\infty}) = \sum_{i=1}^{\infty} P(A_i)$. The number $P(A)$ is called the probability of the event A. The triple $(\Omega, \mathcal{F}, P)$ is called a probability space.

Technical aside: formally $P(A)$ is only defined for $A \in \mathcal{F}$. We will adopt the convention throughout that unless otherwise stated all subsets of interest belong to $\mathcal{F}$.

Proposition A probability measure P satisfies:

(i) $P(A^C) = 1 - P(A)$

(ii) $P(\emptyset) = 0$

(iii) If $A \subseteq B$ then $P(A) \leq P(B)$.

(iv) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof

(i) From II and III, $1 = P(\Omega) = P(A \cup A^C) = P(A) + P(A^C)$, since $A \cap A^C = \emptyset$.

(ii) $P(\emptyset) = P(\Omega^C) = 1 - P(\Omega)$ (by 1) $= 1 - 1$ (by II) $= 0$.

(iii) For $A \subseteq B,\ B = A \cup (B \cap A^C)$ (disjoint). From III, $P(B) = P(A) + P(B \cap A^C) \geq P(A)$, since $P(B \cap A^C) \geq 0$ by I.

(iv) Write $A \cup B = A \cup (B \cap A^C)$ and $B = (B \cap A) \cup (B \cap A^C)$ (disjoint). From III $P(A \cup B) = P(A) + P(B \cap A^C)\,(*)$ and $P(B) = P(B \cap A) + P(B \cap A^C)\,(**)$. (iv) follows by subtracting $(**)$ from (*) and rearranging.

Theorem (Inclusion-Exclusion formula)

$$\begin{aligned} P(\bigcup_{i=1}^{n} A_i) &= \sum_{i=1} nP(A_i) - \sum_{i<j} P(A_i \cap A_j) + \ldots \\ &\quad +(-1)^{r-1} \sum_{i_1<i_2<\ldots<i_r} P(A_{i_1} \cap \ldots \cap A_{i_r}) + \ldots + (-1)^{n-1} P(A_1 \cap \ldots \cap A_n) \end{aligned}$$

The summation $\sum_{i_1<i_2<\ldots<i_r}$ means summation over all $\binom{n}{r}$ sets of indices which are subsets of $1, \ldots, n$ of size r.

Proof by induction
Property (iv) of the previous theorem gives the result for $n = 2$. Further, by (iv) again,

$$\begin{aligned} P(A_1 \cup \ldots \cup A_n) &= P((A_1 \cup \ldots A_{n-1}) \cup A_n) \\ &= P(A_1 \cup \ldots \cup A_{n-1}) + P(A_n) - P((A_1 \cup \ldots \cup A_{n-1}) \cap A_n) \\ &= P(A_1 \cup \ldots \cup A_{n-1}) + P(A_n) - P((A_1 \cap A_n) \cup \ldots \cup (A_{n-1} \cap A_n)) \end{aligned}$$

Now assume the result for $n - 1$ and substitute for first and third terms to give the result.

Example A group of n people place their coats in a pile. Later they each take one coat at random from the pile. What is the probability that at least one person has their own coat?

Take as sample space the $n!$ possible assignments of coats to people. Let A_k be the event that person k has their own coat. To get $P(\bigcup_{i=1}^{n} A_i$ use inclusion-exclusion formula.

Note that

$$\begin{aligned} P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_r} &= \frac{(n-r)!}{n!} \\ \text{Thus} \sum_{i_1<i_2<\ldots<i_r} p(A_{i_1} \cap \ldots A_{i_r} &= \binom{n}{r} \frac{(n-r)!}{n!} = \frac{1}{r!}, \\ \text{and so } P(\bigcup_{i=1}^{n} A_i) &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^(n-1) \frac{1}{n!} \to 1 - e^{-1} \text{ as } n \to \infty \end{aligned}$$

1.5 Conditional probability and independence

Sometimes we may have partial information about the outcome of an experiment. In general this partial information will change the calculation of probabilities. For example, having thrown a fair die, we may be todl that the outcome is even. Then the probability of a 1, 3 or 5 becomes zero, and the probability of a 2, 4 or 6 becomes 1/3.

Definition Provided $P(B) > 0$, define a conditional probability of A given B, written $P(A|B)$, by $P(A|B) = \frac{P(A\cap B)}{P(B)}$.

Remark For fixed B with $P(B) > 0$ we can define a new function Q on $\mathcal{F}$ by $Q(A) \equiv P(A|B)$. It is straightforward to check that Q is a probability measure.

Example You are playing poker. Define $R \equiv$"I have a royal flush", $E \equiv$"My hand contains A♠". Find $P(R|E)$.

1. By definition,
$$P(R|E) = \frac{P(R\cap E)}{P(E)} = \frac{1/\binom{52}{5}}{\binom{51}{4}/\binom{52}{5}} = \frac{1}{\binom{52}{4}}$$

2. From first principles. Assume one card is A♠ and consider new experiment relating only to the unknown values of the other four cards. Each of the $\binom{51}{4}$ outcomes is equally likely and exactly one will result in a royal flush. The required condition probability is thus $1/\binom{51}{4}$.

Note that $P(R) = 4/\binom{52}{5}$, and so $P(R|E) = \frac{13}{5}P(R) > P(R)$. Knowledge that E has occurred changes (here increases) the probability of R occurring.

Example A hat contains three cards. One card is black on both sides, one is black on one side and white on the other, and one is white on both sides. A card is drawn at random and placed on the table. The visible side is black. What is the probability that the other side is black?

Label the faces of the cards b_1, b_2 for black-black, w_1, w_2 for white-white and b_3, w_3 for black-white. Sample space $\Omega = \{(b_1, b_2), (b_2, b_1), (w_1, w_2), (w_2, w_1), (b_3, w_3), (w_3, b_3)\}$ (first number is upper face). All six outcomes equally likely.

Define event $B_U = \{(b_1, b_2), (b_2, b_1), (b_3, w_3)\}$ (black uppermost) and $B_D = \{(b_1, b_2), (b_2, b_1), (w_3, b_3)\}$ (black downermost). Then

$$P(B_D|B_U) = \frac{P(B_D\cap B_U)}{P(B_U)} = \frac{P(\{(b_1, b_2), (b_2, b_1)\})}{P(B_U)} = \frac{2/6}{3/6} = \frac{2}{3}.$$

Theorem (Properties of conditional probability)

1. $P(A\cap B) = P(A)P(B|A) = P(B)P(A|B)$
2. $P(A\cap B\cap C) = P(A)P(B|A)P(C|A\cap B)$
3. $P(A|B\cap C) = \frac{P(A\cap B|C)}{P(B|C)}$

Proof Immediate from definition of conditional probability.

We call events A and B *independent* if the occurrence of one of them does not affect the probability of the other.

Definition Events A and B are independent if $P(A \cap B) = P(A)P(B)$. More generally, a colelction of events $A_i (i \in I)$ are independent if $P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i)$ for all finite subsets J of I.

Note that A and B independent $\Rightarrow P(A|B) = P(A)$ and $P(B|A) = P(B)$.

Example Two dice are thrown. The sample space $\Omega = (i, j) : 1 \leq i, j \leq 6$ has 36 equally likely outcomes. Let $A_1 = \{\text{first die odd}\}$, $A_2 = \{\text{second die odd}\}$, $A_3 = \{\text{ sum is odd}\}$. Are A_1, A_2 independent? $P(A_1) = P(A_2) = 18/36$, $P(A_1 \cap A_2) = 9/36$. Then $P(A_1 \cap A_2) = P(A_1)P(A_2)$, so yes.

Similarly A_2, A_3 and A_1, A_3 are independent so the three events are pairwise independent.

Are A_1, A_2, A_3 independent? $P(A_1 \cap A_2 \cap A_3) = 0 \neq 1/8 = P(A_1)P(A_2)P(A_3)$, so no.

Example You are asked a series of questions and you get each right with probability p. The outcomes of different questions are independent (A sequence of Bernoulli trials). The probability that the first correct answer is at the rth question is $p_r = (1-p)^{r-1}p$. Since $\sum_{r=1}^{\infty} p_r = 1$ you've got to get a question right eventually.

Theorem (Law of Total Probability / Partition Theorem) Let $B_1, B_2, \ldots, B_n$ be a partition of Ω. Then $P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$.

Proof If $i \neq j$ then $A \cap B_i$ and $A \cap B_j$ are disjoint. Then

$$\sum_{i=1}^{n} P(A|B_i)P(B_i) = \sum P(A \cap B_i) = P(\bigcup_{i=1}^{n} (A \cap B_i)) = P(A)$$

Example The probability I remember to lock my bike is 19/20. If I forget to lock it, it's stolen with probability 4/5, while if I do lock it it's stolen with probability 1/100. What is the probability that it's stolen tomorrow?

Let $L = \{\text{bike locked}\}$, $S = \{\text{bike stolen}\}$. Then $P(L) = 19/20$ and $P(L^C) = 1/20$. $P(S|L) = 1/100$, $P(S|L^C) = 4/5$. Now $P(S) = P(S|L)P(L) + P(S|L^C)P(L^C) = 99/2000 = 0.0495$ by above theorem using the partition L, L^C.

Theorem (Bayes Formula) Let $B_1, B_2, \ldots, B_n$ be a partition of Ω. Then

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$$

Proof

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$$

(top half by definition of conditional probability, bottom by previous result)

Example A diseae occurs by chance in one in every 200 people. A random person is tested; if they have the disease, the test will correctly say so with probability 0.95; if not, the test will wrongly say they do with probability 0.01. Find the probability that (i) The test says a person has the disease (ii) the person has the disease given that the test says they do.

Let $D = \{\text{person has disease}\}$, $A = \{\text{test says they do}\}$.

(i $P(A) = P(A|D)P(D) + P(A|D^C)P(D^C)$ (using partition D, D^C) $= 0.95 \times (1/200) + 0.01 \times (199/200) = 0.0147$.

(ii By Bayes formula, $P(D|A) = \frac{(P(A|D)P(D)}{0.0147} = \frac{0.95\times(1/200)}{0.0147} = 95/294 \approx 0.32$.

Suppose instead that the person is being examined by a doctor because they are feeling ill. After the examination but before the test, the doctor reckons $P(D) = 0.3$. Then $P(A) = 0.95 \times (3/10) + 0.01 \times (7/10) = 0.292$ and $P(D|A) = \frac{P(A|D)P(D)}{0.292} = \frac{0.95\times(3/10)}{0.292} = 285/292 \approx 0.976$.

2 Discrete Random Variables

Example

A roulette wheel has 38 slots: 18 red, 18 black, 2 green. Suppose the red slots are labelled 1,3,...,35, the black slots 2,4,...,36 and the green 0 and 00. A gambler bets £1 on red. She wins £1 for red and loses otherwise. So $\Omega = \{00, 0, 1, \ldots, 36\}$ and if the outcome is ω, the gambler wins $X(\omega)$ where $X(1) = X(3) = \ldots = X(35) = 1$, $X(00) = X(0) = X(2) = \ldots = X(36) = -1$. The function $X : \Omega \to \mathbf{R}$ is a *discrete random variable.*

Take Ω to be finite or countable, with P defined on all subsets of Ω.

Definition A discrete random variable S is a real-valued function defined on the sample space Ω.

Technical aside If we do not assume that Ω is countable then $X : \Omega \to \mathbf{R}$ is a discrete random variable on the probability space $(\Omega, \mathcal{F}, P)$ if

1. $\{X(\omega) : \omega \in \Omega\}$ is a countable set
2. $\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F} \forall x \in \mathbf{R}$

(1) ensures X is discrete (takes a countable set of values). (2) needed to ensure that the event "X takes value X" belongs to event space $\mathcal{F}$.

Example

Suppose you throw two dice: $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$. We can define random variables Y and Z by $Y(i, j) = i + j$ and $Z(i, j) = \max(i, j)$.

Write R_X for the range of X, i.e. the set of values that the random variable X can take.

Let $R_X = \{x_i : i = 1, 2, \ldots\}$ (finite or countable since Ω is). Write $P(X = x_i) = P(\{X = x_i\}) = P(\{\omega \in \Omega : X(\omega) = x_i\})$

$$= \sum_{\omega:X(\omega)=x_i} P(\{\omega\})$$

Further, for $A \subset \mathbf{R}$, $P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\})$

$$= \sum_{\omega:X(\omega)\in A} P(\{\omega\}) = \sum_{x_i\in A} \sum_{\omega:X(\omega)=x_i} P(\{\omega\}) = \sum_{x_i\in A} P(X = x_i)$$

Examples

In the roulette example, $R_X = \{-1, 1\}$. $P(X = 1) = p_1 + p_2 + \ldots + p_{35} = 18/38 = 9/19$. $p_i = P\{\text{outcome is } i\}$. $P(X = -1) = p_{00} + p_0 + p_2 + \ldots + p_{36} = 20/38 = 10/19$.

Dice example: $R_Y = \{2, 3, \ldots, 12\}.P(Y \geq 11) = P(Y = 11) + P(Y = 12) = 2/36 + 1/36 = 1/12$.

Definition The probabilities $P(X = x_i)$, $x_i \in R_X$ are referred to as the *(probability) distribution* of the random variable X. The function $p_X : R_X \to [0, 1]$ defined by $p_X(X_i) = P(X = x_i)$ is called the *probability mass function* of X. Sometimes we abbreviate $p_X(x)$ to $p(x)$. Note that since $P(\Omega) = 1$, $\sum_{x \in R_X} p_X(x) = 1$.

For roulette example, $p(-1) = 10/19$, $p(1) = 9/19$.

Examples

1. A random variable X with $R_X = \{0, 1\}$ is said to have a Bernoulli distribution if $P(X = 1) = p$, $P(X = 0)1 - p$. e.g. conduct an experiment with two outcomes called "success" and "failure" and let p be the probability of success. Associate $X = 1$ with success, $X = 0$ with failure.

2. A random variable X with $R_X = \{0, \ldots, n\}$ is said to have a Binonmial distribution with parameters n and p if $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ $k = 0, \ldots, n$. e.g. X is no. of successes in n independent trials of the experiment in (1).

3. A random variable X with $R_X = \{1, 2, \ldots\}$ is said to have a Geometric distribution with parameter p if $P(X = k) = (1 - p)^{k-1} p \quad k = 1, 2, \ldots$. e.g. repeat independent trials of experiment in (1) and define X to be number of trials required to first get a success.

4. Suppose $\lambda > 0$ is fixed. A random variable X with $R_X = \{0, 1, \ldots\}$ is said to have a Poisson distribution with parameter λ if $P(X = k) = (e^{-\lambda} \lambda^k)/k!$ for $k = 0, 1, 2, \ldots$.

   It turns out that the Poisson distribution provides a good description of the numbers of "rare" events over some time period, e.g. no. of fatal accidents in a region in a year or number of paricles emitted by a radioactive sauce.

2.1 Expectation

Definition The *expectation* (or mean) of a random variable X is the number $E(X) = \sum_{x \in R_X} x P(X = x) = \sum_{x \in R_X} x p_x(x)$ provided the sum converges absolutely. (i.e. provided $\sum_{x \in R_X} |x P(X = x)| < \infty$)

Examples

1. Bernoulli (p) distribution: $P(X = 0) = (1-p)$, $P(X = 1) = p$, $E(X) = 0.(1-p) + 1.p = p$.

2. Poisson (λ) distribution.

$$E(X) = \sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!}$$

$$\text{(Put } x-1=k) \qquad = e^{-\lambda}\lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda}\lambda e^{\lambda} = \lambda$$

2.2 Functions of random variables

If X is a discrete random variable and $g : R_X \to \mathbf{R}$ is any function, then $Y = g(X)$ is also a random variable defined by $Y(\omega) = g(X(\omega))$ for $\omega \in \Omega$.

e.g. for constants a, b, c, $Y_1 = aX + b$, $Y_2 = (X - c)^2$ are random variables taking values $aX(\omega) + b$, $(X(\omega) - c)^2$, $\omega \in \Omega$. In general, for $Y = g(X)$, $P(Y = y) = P(g(X) = y) = \sum_{x g(x)=y} P(X = x)$. The expectation of the random variable $g(X)$ is

$$E(g(X)) = \sum_{y\in R_Y} yP(g(x) = y) = \sum_{y\in R_Y} y \sum_{x g(x)=y} P(X = x) = \sum_{x\in R_X} g(x)P(X = x)$$

Theorem (Properties of Expectation) Suppose a and b are constants.

1. If $X \geq a$ (i.e. if $X(\omega) \geq a \forall \omega \in \Omega$) then $E(X) \geq a$.
2. If $P(X = b) = 1$ then $E(X) = b$.
3. $E(aX + b) = aE(X) + b$
4. $E(g(X) + h(X)) = E(g(X)) + E(h(X))$

Proof

1. $E(X) = \sum_{x\in R_X} xP(X = x) \geq \sum_{x\in R_X} aP(X = x) = a$
2. $P(X = x) = \begin{cases} 1 & : \text{ if } x = b \text{ so } E(X) = 1.b + 0 = b \\ 0 & : \text{ otherwise} \end{cases}$
3. $$E(aX + b) = \sum_{x\in R_X} (ax + b)P(X = x) =$$
$$a \sum_{x\in R_X} xP(X = x) + b \sum_{x\in R_X} P(X = x) = aE(X) + b$$
4. $$E(g(X) + h(X)) = \sum_{x\in R_X} (g(x) + h(x))P(X = x)$$
$$= \sum_{x\in R_X} g(x)P(X = x) + \sum_{x\in R_X} h(x)P(X = x) = E(g(x)) + E(h(X))$$

Definition The variance of a random variable X, usually written $\text{Var}(X)$, is defined by $\text{Var}(X) = E((X - E(X))^2)$. So writing $\mu = E(X)$, $\text{Var}(X) = E((X - \mu)^2) = \sum_{x\in R_X} (x - \mu)^2 P(X = x)$.

The variance of a random variable means how "spread out" or dispersed its distribution is around its mean.

Example

Suppose $X = \begin{cases} -1 & : \text{ with prob. } 1/2 \\ 1 & : \text{ with prob. } 1/2 \end{cases}$ and $Y = \begin{cases} -100 & : \text{ with prob. } 1/2 \\ 100 & : \text{ with prob. } 1/2 \end{cases}$. Then $E(X) = E(Y) = 0$. $\text{Var}(X) = \frac{1}{2}(-1-0)^2 + \frac{1}{2}(1-0)^2 = 1$, $\text{Var}(Y) = \frac{1}{2}(-100-0)^2 + \frac{1}{2}(100-0)^2 = 10000$.

Theorem (Properties of Variance)

1. $\text{Var}(X) \geq 0$.
2. If a, b are constants then $\text{Var}(aX+b) = a^2\text{Var}(X)$.
3. $\text{Var}(X) = E(X^2) - [E(X)]^2$

Proof

1. $\text{Var}(X) = \sum_{x \in R_X} (x-\mu)^2 P(X=x) \geq \sum_{x \in R_X} 0.P(X=x) = 0$.
2. $\text{Var}(aX+b) = E[(aX+b-E(aX+b))^2] = E[(aX+b-aE(X)-b)^2] = E[(a(X-E(X)))^2] = a^2\text{Var}(X)$
3. $\text{Var}(X) = E((X-\mu)^2) = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu\mu + \mu^2 = E(X^2) - \mu^2$

Examples

1. Bernoulli (p) distribution. $P(X=1) = p = 1 - P(X=0)$. $E(X^2) = \sum_{x \in R_X} x^2 P(X=x) = 1^2.p + 0^2(1-p) = p$. $\text{Var}(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p)$.
2. Poisson (λ) distribution: $P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $x = 0, 1, \ldots$

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 \frac{e^{-\lambda}\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda} \left[\sum_{k=1}^{\infty} (k-1) \frac{\lambda^(k-1)}{(k-1)!} + \sum_{k=1}^{\infty} 1. \frac{\lambda^{k-1}}{(k-1)!} \right] \\ &= \lambda e^{-\lambda} \left[\lambda \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right] \end{aligned}$$

$= \lambda e^{-\lambda}(\lambda e^{\lambda} + e^{\lambda}) = \lambda^{\lambda}$