Introduction to Probability

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 308

Subsection 1

Random Variables

Random Variables

- When an experiment is performed, we may be interested in some function of the outcome as opposed to the actual outcome itself.

  Example: In tossing dice, we may be interested in the sum of the two dice and not concerned about the separate values of each die.

  That is, we may be interested in knowing that the sum is 7 and may not be concerned over whether the actual outcome was (1, 6), (2, 5), (3, 4), (4, 3), (5, 2) or (6, 1).

  Example: In flipping a coin, we may be interested in the total number of heads that occur and not care at all about the actual head-tail sequence that results.
- These real valued functions defined on the sample space, are known as **random variables**.
- Because the value of a random variable is determined by the outcome of the experiment, we may assign probabilities to the possible values of the random variable.

Example

- Suppose that our experiment consists of tossing 3 fair coins.

  If we let Y denote the number of heads that appear, then Y is a random variable taking on one of the values 0, 1, 2 and 3 with respective probabilities

$$\begin{array}{rcl} P\{Y=0\} & = & P\{(T,T,T)\} = \frac{1}{8}; \\ P\{Y=1\} & = & P\{(T,T,H),(T,H,T),(H,T,T)\} = \frac{3}{8}; \\ P\{Y=2\} & = & P\{(T,H,H),(H,T,H),(H,H,T)\} = \frac{3}{8}; \\ P\{Y=3\} & = & P\{(H,H,H)\} = \frac{1}{8}. \end{array}$$

  Since Y must take on one of the values 0 through 3, we must have

$$1 = P(\bigcup_{i=0}^{3}\{Y=i\}) = \sum_{i=0}^{3} P\{Y=i\},$$

  which, of course, is in accord with the preceding probabilities.

Example

- Three balls are to be randomly selected without replacement from an urn containing 20 balls numbered 1 through 20.

  If we bet that at least one of the balls that are drawn has a number as large as or larger than 17, what is the probability that we win the bet?

  Let X denote the largest number selected.

  Then X is a random variable taking on one of the values $3, 4, \ldots, 20$.

  Assume that each of the $\binom{20}{3}$ possible selections are equally likely.

  Then
  $$P\{X = i\} = \frac{\binom{i-1}{2}}{\binom{20}{3}}, \quad i = 3, \ldots, 20.$$
  This follows because the number of selections that result in the event $\{X = i\}$ is the number of selections that result in the ball i and two of the balls 1 through $i-1$ being chosen.

Example (Cont'd)

- Now we compute:

$$P\{X=20\}=\frac{\binom{19}{2}}{\binom{20}{3}}=\frac{3}{20};\quad P\{X=19\}=\frac{\binom{18}{2}}{\binom{20}{3}}=\frac{51}{380};$$
$$P\{X=18\}=\frac{\binom{17}{2}}{\binom{20}{3}}=\frac{34}{285};\quad P\{X=17\}=\frac{\binom{16}{2}}{\binom{20}{3}}=\frac{2}{19}.$$

  The event $\{X\geq 17\}$ is the union of the disjoint events $\{X=i\}$, $i=17,18,19,20$.

  Thus, the probability of our winning the bet is

$$P\{X\geq 17\}=\frac{3}{20}+\frac{51}{380}+\frac{34}{285}+\frac{2}{19}.$$

Example

- Independent trials consisting of the flipping of a coin having probability p of coming up heads are continually performed until either a head occurs or a total of n flips is made.
  If we let X denote the number of times the coin is flipped, then X is a random variable taking on one of the values $1, 2, 3, \ldots, n$. Then:

$$\begin{array}{rcl} P\{X=1\} & = & P\{H\} = p; \\ P\{X=2\} & = & P\{(T,H)\} = (1-p)p; \\ P\{X=3\} & = & P\{(T,T,H)\} = (1-p)^2 p; \\ & \vdots & \\ P\{X=n-1\} & = & P\{(\underbrace{T,T,\ldots,T}_{n-2},H)\} = (1-p)^{n-2}p; \\ P\{X=n\} & = & P\{(\underbrace{T,T,\ldots,T}_{n-1},T),(\underbrace{T,T,\ldots,T}_{n-1},H)\} \\ & = & (1-p)^{n-1}. \end{array}$$

Example

- Three balls are randomly chosen from an urn containing 3 white, 3 red and 5 black balls.
  - We win \$1 for each white ball selected;
  - We lose \$1 for each red ball selected.

  Let X denote our total winnings from the experiment.

  X is a random variable taking on the possible values $0, \pm 1, \pm 2, \pm 3$ with respective probabilities:

$$\begin{aligned}
P\{X=0\} &= \frac{\binom{5}{3}+\binom{3}{1}\binom{3}{1}\binom{5}{1}}{\binom{11}{3}} = \frac{55}{165};\\
P\{X=1\} &= P\{X=-1\} = \frac{\binom{3}{1}\binom{5}{2}+\binom{3}{2}\binom{3}{1}}{\binom{11}{3}} = \frac{39}{165};\\
P\{X=2\} &= P\{X=-2\} = \frac{\binom{3}{2}\binom{5}{1}}{\binom{11}{3}} = \frac{15}{165};\\
P\{X=3\} &= P\{X=-3\} = \frac{\binom{3}{3}}{\binom{11}{3}} = \frac{1}{165}.
\end{aligned}$$

Example (Cont'd)

- These probabilities are obtained, for instance, by noting that:
  - In order for X to equal 0, either all 3 balls selected must be black or 1 ball of each color must be selected;
  - The event $\{X = 1\}$ occurs either if 1 white and 2 black balls are selected or if 2 white and 1 red is selected.

  $\vdots$

  The probability that we win money is given by

$$\sum_{i=1}^{3} P\{X = i\} = \frac{55}{165} = \frac{1}{3}.$$

Example

- Suppose that there are N distinct types of coupons and that each time one obtains a coupon, it is, independently of previous selections, equally likely to be any one of the N types.

  Let T be the number of coupons that needs to be collected until one obtains a complete set of at least one of each type.

  Rather than derive $P\{T=n\}$ directly, let us start by considering the probability that T is greater than n.

  To do so, fix n and define the events $A_1, A_2, \ldots, A_N$ as follows:

  A_j is the event that no type j coupon is contained among the first n coupons collected, $j = 1, \ldots, N$.

  Hence,

$$\begin{aligned} P\{T > n\} &= P(\bigcup_{j=1}^N A_j) \\ &= \sum_j P(A_j) - \sum_{j_1<j_2} P(A_{j_1}A_{j_2}) + \cdots \\ &\quad + (-1)^{k+1} \sum_{j_1<j_2<\cdots<j_k} P(A_{j_1}A_{j_2}\cdots A_{j_k}) + \cdots \\ &\quad + (-1)^{N+1} P(A_1A_2\cdots A_N). \end{aligned}$$

Example (Cont'd)

- Now, A_j will occur if each of the n coupons collected is not of type j. Each of the coupons will not be of type j with probability $\frac{N-1}{N}$.
  By the assumed independence of the types of successive coupons, $P(A_j) = (\frac{N-1}{N})^n$.
  Also, the event $A_{j_1}A_{j_2}$ will occur if none of the first n coupons collected is of either type j_1 or type j_2.
  Again using independence, we see that $P(A_{j_1}A_{j_2}) = (\frac{N-2}{N})^n$.
  The same reasoning gives $P(A_{j_1}A_{j_2}\cdots A_{j_k}) = (\frac{N-k}{N})^n$.
  We see that, for $n > 0$,
  $$\begin{array}{rcl} P\{T > n\} & = & N(\frac{N-1}{N})^n - \binom{N}{2}(\frac{N-2}{N})^n + \binom{N}{3}(\frac{N-3}{N})^n - \cdots \\ & & + (-1)^N\binom{N}{N-1}(\frac{1}{N})^n \\ & = & \sum_{i=1}^{N-1}\binom{N}{i}(\frac{N-i}{N})^n(-1)^{i+1}. \end{array}$$
  The probability that T equals n can now be obtained by using
  $$P\{T = n\} = P\{T > n-1\} - P\{T > n\}.$$

Example

- Another random variable of interest is the number D_n of distinct types of coupons that are contained in the first n selections.

  To compute $P\{D_n = k\}$, we start by fixing attention on a particular set of k distinct types.

  We determine the probability that this set constitutes the set of distinct types obtained in the first n selections.

  In order for this to be the situation, it is necessary and sufficient that, of the first n coupons obtained:

  A: Each is one of these k types;
  B: Each of these k types is represented.

  Each coupon selected will be one of the k types with probability $\frac{k}{N}$.

  So the probability that A will be valid is $(\frac{k}{N})^n$.

Example (Cont'd)

- Also, given that a coupon is of one of the (fixed) k types, it is easy to see that it is equally likely to be of any one of these k types.

  Hence, the conditional probability of B given that A occurs is the same as the probability that a set of n coupons, each equally likely to be any of k possible types, contains a complete set of all k types.

  This is the probability that the number needed to amass a complete set, when choosing among k types, is less than or equal to n.

  So, it is obtainable from

$$\sum_{i=1}^{k-1} \binom{k}{i} \left(\frac{k-i}{k}\right)^n (-1)^{i+1}.$$

Example (Cont'd)

- We obtained:

$$\begin{array}{rcl} P(A) & = & (\frac{k}{N})^n; \\ P(B|A) & = & 1 - \sum_{i=1}^{k-1} \binom{k}{i} (\frac{k-i}{k})^n (-1)^{i+1}. \end{array}$$

Moreover, there are $\binom{N}{k}$ possible choices for the set of k types. Hence, we arrive at

$$\begin{array}{rcl} P\{D_n = k\} & = & \binom{N}{k} P(AB) \\ & = & \binom{N}{k} (\frac{k}{N})^n [1 - \sum_{i=1}^{k-1} \binom{k}{i} (\frac{k-i}{k})^n (-1)^{i+1}]. \end{array}$$

Cumulative Distribution Function

- For a random variable X, the function F defined by

$$F(x) = P\{X \leq x\}, \quad -\infty < x < \infty,$$

is called the **cumulative distribution function**, or, more simply, the **distribution function**, of X.

- Thus, the distribution function specifies, for all real values x, the probability that the random variable is less than or equal to x.
- Now, suppose that $a \leq b$.

  Note that the event $\{X \leq a\}$ is contained in the event $\{X \leq b\}$.

  Thus, $F(a)$, the probability of the former, is less than or equal to $F(b)$, the probability of the latter.

  In other words, $F(x)$ is a nondecreasing function of x.

Subsection 2

Discrete Random Variables

Discrete Random Variables and Probability Mass

- A random variable that can take on at most a countable number of possible values is said to be **discrete**.
- For a discrete random variable X, we define the **probability mass function** $p(a)$ of X by

$$p(a) = P\{X = a\}.$$

- The probability mass function $p(a)$ is positive for at most a countable number of values of a.
- That is, if X must assume one of the values $x_1, x_2, \ldots$, then

$$\begin{array}{ll} p(x_i) \geq 0, & \text{for } i = 1, 2, \ldots, \\ p(x) = 0, & \text{for all other values of } x. \end{array}$$

- Since X must take on one of the values x_i, we have $\sum_{i=1}^{\infty} p(x_i) = 1$.

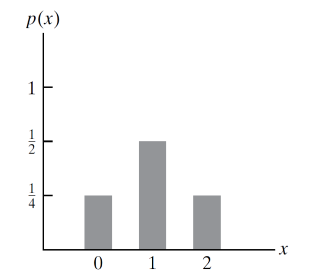
Graphical Representation of Probability Mass

- The probability mass function is often presented in a graphical format by plotting $p(x_i)$ on the y-axis against x_i on the x-axis.

  Example: If the probability mass function of X is

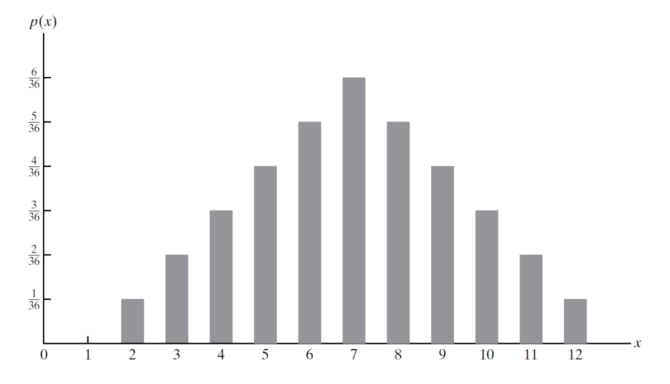
  $$p(0) = \frac{1}{4},\ p(1) = \frac{1}{2},\ p(2) = \frac{1}{4},$$

  we can represent this function graphically as in the diagram:



Another Example

- Similarly, a graph of the probability mass function of the random variable representing the sum when two dice are rolled is depicted below:



Example

- The probability mass function of a random variable X is given by $p(i) = \frac{c\lambda^i}{i!}$, $i = 0, 1, 2, \ldots$, where λ is some positive value. Find:
  (a) $P\{X = 0\}$;
  (b) $P\{X > 2\}$.

  First, we determine c:

$$\sum_{i=0}^{\infty} p(i) = 1 \;\Rightarrow\; c \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = 1$$

$$\overset{\sum_{i=0}^{\infty} \frac{x^i}{i!} = e^x}{\Longrightarrow} \quad ce^{\lambda} = 1 \;\Rightarrow\; c = e^{-\lambda}.$$

(a) $P\{X = 0\} = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda}$.

(b)

$$\begin{aligned} P\{X > 2\} &= 1 - P\{X \leq 2\} \\ &= 1 - P\{X = 0\} - P\{X = 1\} - P\{X = 2\} \\ &= 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2}. \end{aligned}$$

Probability Mass and Cumulative Distribution

- The **cumulative distribution function** F can be expressed in terms of $p(a)$ by
$$F(a) = \sum_{\text{all } x \leq a} p(x).$$
- If X is a discrete random variable whose possible values are $x_1, x_2, x_3, \ldots$, where $x_1 < x_2 < x_3 < \cdots$, then the distribution function F of X is a *step function*.

  That is:
  - The value of F is constant in the intervals $[x_{i-1}, x_i)$;
  - It then takes a step (or jump) of size $p(x_i)$ at x_i.

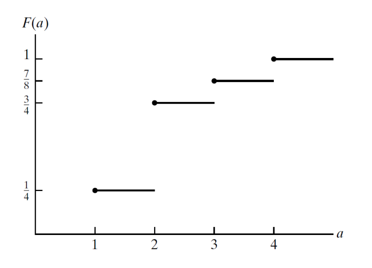
Example

- Suppose X has a probability mass function given by

$$p(1) = \frac{1}{4}, \quad p(2) = \frac{1}{2}, \quad p(3) = \frac{1}{8}, \quad p(4) = \frac{1}{8}.$$

Then its cumulative distribution function is

$$F(a) = \begin{cases} 0, & \text{if } a < 1 \\ \frac{1}{4}, & \text{if } 1 \le a < 2 \\ \frac{3}{4}, & \text{if } 2 \le a < 3 \\ \frac{7}{8}, & \text{if } 3 \le a < 4 \\ 1, & \text{if } 4 \le a. \end{cases}$$



Note that the size of the step at any of the values 1, 2, 3 and 4 is equal to the probability that X assumes that particular value.

Subsection 3

Expected Value

Expected Value

- If X is a discrete random variable having a probability mass function $p(x)$, then the **expectation**, or the **expected value**, of X, denoted by $E[X]$, is defined by

$$E[X] = \sum_{x:p(x)>0} xp(x).$$

- In words, the expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes it.

Example

- Suppose the probability mass function of X is given by

$$p(0) = \frac{1}{2} = p(1).$$

  Then

$$E[X] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

  It is just the ordinary average of the two possible values, 0 and 1, that X can assume.

- Suppose, on the other hand, that $p(0) = \frac{1}{3}$, $p(1) = \frac{2}{3}$,

  Then

$$E[X] = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}.$$

  This is a weighted average of the two possible values 0 and 1, where (since $p(1) = 2p(0)$) value 1 is given twice as much weight as value 0.

Example

- Let X be the outcome when we roll a fair die.
  Find $E[X]$.
  Since the die is fair, we get
  $$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}.$$
  Thus, we obtain
  $$\begin{aligned} E[X] &= 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} \\ &\quad + 4 \cdot \tfrac{1}{6} + 5 \cdot \tfrac{1}{6} + 6 \cdot \tfrac{1}{6} \\ &= \tfrac{7}{2}. \end{aligned}$$

The Indicator Variable of an Event

- We say that I is an **indicator variable** for the event A if

$$I = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{if } A^c \text{ occurs.} \end{cases}$$

Find $E[I]$.

By definition,

$$p(1) = P(A), \quad p(0) = 1 - P(A).$$

Hence, we have

$$E[I] = P(A).$$

That is, the expected value of the indicator variable for the event A is equal to the probability that A occurs.

Example

- A contestant on a quiz show is presented with two questions, Question 1 and Question 2.

  He is to attempt to answer them in some order he chooses.
  - If he decides to try Question i first, then he will be allowed to go on to question j, $j \neq i$, only if his answer to question i is correct.
  - If his initial answer is incorrect, he is not allowed to answer the other question.

  The contestant is to receive V_i dollars if he answers question i correctly, $i = 1, 2$.

  E.g., he will receive $V_1 + V_2$ dollars if he answers both questions correctly.

  Suppose the probability that he knows the answer to Question i is P_i.

  Which question should he attempt to answer first so as to maximize his expected winnings?

Example (Cont'd)

- Assume that the events E_i, $i = 1, 2$, that he knows the answer to question i are independent events.
  - If he attempts to answer Question 1 first, then he will win
    - 0 with probability $1 - P_1$;
    - V_1 with probability $P_1(1 - P_2)$;
    - $V_1 + V_2$ with probability $P_1 P_2$.

    Hence, his expected winnings in this case will be $V_1 P_1(1 - P_2) + (V_1 + V_2)P_1 P_2$.
  - If he attempts to answer Question 2 first, his expected winnings will be (by symmetry), $V_2 P_2(1 - P_1) + (V_1 + V_2)P_1 P_2$.

  Therefore, it is better to try question 1 first if

$$V_1 P_1(1 - P_2) \geq V_2 P_2(1 - P_1).$$

  This is equivalent to

$$\frac{V_1 P_1}{1 - P_1} \geq \frac{V_2 P_2}{1 - P_2}.$$

Example

- A school class of 120 students is driven in 3 buses to a theater. There are 36 students in one bus, 40 in another, and 44 in the third. When the buses arrive, one of the 120 students is randomly chosen. Let X be the number of students on the bus of the chosen student. Find $E[X]$.

  The randomly chosen student is equally likely to be any of the 120. Hence,

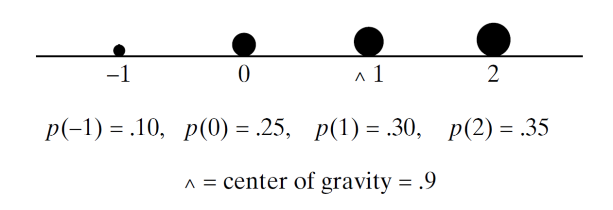
  $$P\{X=36\}=\frac{36}{120},\ P\{X=40\}=\frac{40}{120},\ P\{X=44\}=\frac{44}{120}.$$

  Hence,

  $$E[X]=36\cdot\frac{3}{10}+40\cdot\frac{1}{3}+44\cdot\frac{11}{30}=\frac{1208}{30}.$$

Expectation and Center of Gravity

- The probability concept of expectation is analogous to the physical concept of the *center of gravity* of a distribution of mass.
- Consider a discrete random variable X having probability mass function $p(x_i)$, $i \geq 1$.
  - On a weightless rod, weights with mass $p(x_i)$, $i \geq 1$, are located at the points x_i, $i \geq 1$;
  - Then the point at which the rod would be in balance is known as the **center of gravity**.



- By using elementary statics, one shows that this point is at $E[X]$.

Subsection 4

Expectation of a Function of a Random Variable

Expected Value of a Function of a Random Variable

- Suppose that we are given a discrete random variable along with its probability mass function and that we want to compute the expected value of some function of X, say, $g(X)$.
- One way to accomplish this is as follows:
    - Since $g(X)$ is itself a discrete random variable, it has a probability mass function.
    - This can be determined from the probability mass function of X.
    - Once it has been determined, we can compute $E[g(X)]$ by using the definition of expected value.

Example

- Let X denote a random variable that takes on any of the values $-1, 0$ and 1 with respective probabilities

$$P\{X = -1\} = 0.2, \quad P\{X = 0\} = 0.5, \quad P\{X = 1\} = 0.3.$$

  Compute $E[X^2]$.

  Let $Y = X^2$.

  Then the probability mass function of Y is given by:

$$\begin{array}{rcl} P\{Y = 1\} & = & P\{X = -1\} + P\{X = 1\} = 0.5; \\ P\{Y = 0\} & = & P\{X = 0\} = 0.5. \end{array}$$

  Hence,

$$E[X^2] = E[Y] = 1 \cdot 0.5 + 0 \cdot 0.5 = 0.5.$$

- Note that $0.5 = E[X^2] \neq (E[X])^2 = 0.1^2 = 0.01$.

Computing the Expectation

Proposition

If X is a discrete random variable that takes on one of the values x_i, $i \geq 1$, with respective probabilities $p(x_i)$, then, for any real-valued function g,

$$E[g(X)] = \sum_i g(x_i)p(x_i).$$

Suppose that y_j, $j \geq 1$, represent the different values of $g(x_i)$, $i \geq 1$. Then, grouping all the $g(x_i)$ having the same value gives

$$\begin{aligned}
\textstyle\sum_i g(x_i)p(x_i) &= \textstyle\sum_j \sum_{i:g(x_i)=y_j} g(x_i)p(x_i) \\
&= \textstyle\sum_j \sum_{i:g(x_i)=y_j} y_j p(x_i) \\
&= \textstyle\sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i) \\
&= \textstyle\sum_j y_j P\{g(X) = y_j\} \\
&= E[g(X)].
\end{aligned}$$

Example (Revisited)

- Let X denote a random variable that takes on any of the values $-1, 0$ and 1 with respective probabilities

$$P\{X=-1\}=0.2, \quad P\{X=0\}=0.5, \quad P\{X=1\}=0.3.$$

Compute $E[X^2]$.

We apply the Proposition:

$$\begin{aligned} E\{X^2\} &= (-1)^2 \cdot 0.2 + 0^2 \cdot 0.5 + 1^2 \cdot 0.3 \\ &= 1 \cdot (0.2+0.3) + 0 \cdot 0.5 \\ &= 0.5. \end{aligned}$$

Example

- A product that is sold seasonally yields a net profit of b dollars for each unit sold and a net loss of ℓ dollars for each unit left unsold when the season ends.

  The number of units of the product that are ordered at a specific department store during any season is a random variable having probability mass function $p(i)$, $i \geq 0$.

  If the store must stock this product in advance, determine the number of units the store should stock so as to maximize its expected profit.

  Let X denote the number of units ordered.

  If s units are stocked, then the profit $P(s)$ can be expressed as

$$P(s) = \begin{cases} bX - (s - X)\ell, & \text{if } X \leq s \\ sb, & \text{if } X > s. \end{cases}$$

Example (Cont'd)

- Hence, the expected profit equals

$$\begin{array}{rcl} E[P(s)] & = & \sum_{i=0}^{s}[bi-(s-i)\ell]p(i)+\sum_{i=s+1}^{\infty}sbp(i) \\ & = & (b+\ell)\sum_{i=0}^{s}ip(i)-s\ell\sum_{i=0}^{s}p(i)+sb[1-\sum_{i=0}^{s}p(i)] \\ & = & (b+\ell)\sum_{i=0}^{s}ip(i)-(b+\ell)s\sum_{i=0}^{s}p(i)+sb \\ & = & sb+(b+\ell)\sum_{i=0}^{s}(i-s)p(i). \end{array}$$

To determine the optimum value of s, let us investigate what happens to the profit when we increase s by 1 unit.

By substitution, we see that the expected profit in this case is given by

$$\begin{array}{rcl} E[P(s+1)] & = & b(s+1)+(b+\ell)\sum_{i=0}^{s+1}(i-s-1)p(i) \\ & = & b(s+1)+(b+\ell)\sum_{i=0}^{s}(i-s-1)p(i). \end{array}$$

Therefore,

$$E[P(s+1)]-E[P(s)]=b-(b+\ell)\sum_{i=0}^{s}p(i).$$

Example (Cont'd)

- We found $E[P(s+1)] - E[P(s)] = b - (b+\ell)\sum_{i=0}^{s} p(i)$.

  Thus, stocking $s+1$ units will be better than stocking s units whenever
  $$\sum_{i=0}^{s} p(i) < \frac{b}{b+\ell}.$$

  Because the left-hand side is increasing in s while the right-hand side is constant, the inequality will be satisfied for all values of $s \leq s^*$, where s^* is the largest value of s satisfying the equation.

  Note that
  $$E[P(0)] < \cdots < E[P(s^*)] < E[P(s^*+1)] > E[P(s^*+2)] > \cdots.$$

  Hence, stocking s^*+1 items will lead to a maximum expected profit.

Linearity of Expectation

Corollary

If a and b are constants, then

$$E[aX+b] = aE[X]+b.$$

$$\begin{aligned} E[aX+b] &= \textstyle\sum_{x:p(x)>0}(ax+b)p(x) \\ &= a\textstyle\sum_{x:p(x)>0} xp(x) + b\textstyle\sum_{x:p(x)>0} p(x) \\ &= aE[X]+b. \end{aligned}$$

- The expected value of a random variable X, $E[X]$, is also referred to as the **mean** or the **first moment** of X.
- The quantity $E[X^n]$, $n \geq 1$, is called the n**th moment** of X.
- By the preceding proposition

$$E[X^n] = \sum_{x:p(x)>0} x^n p(x).$$

Subsection 5

Variance

Measuring the Variation of a Distribution

- Let X be a random variable.
- Let F be its distribution function.
- $E[X]$ yields the weighted average of the possible values of X, but it does not tell us anything about the variation, or spread, of the values.

  Example: Consider random variables W, Y and Z having probability mass functions determined by:

$$W = 0 \text{ with probability } 1,$$

$$Y = \begin{cases} -1, & \text{with probability } \frac{1}{2} \\ +1, & \text{with probability } \frac{1}{2} \end{cases}, \quad Z = \begin{cases} -100, & \text{with probability } \frac{1}{2} \\ +100, & \text{with probability } \frac{1}{2} \end{cases}$$

  All of W, Y, Z have the same expectation 0.

  But there is a much greater spread in the possible values of Y than in those of W (which is a constant) and in the possible values of Z than in those of Y.

The Variance

- We expect X to take on values around its mean $E[X]$.
- So a reasonable way of measuring the possible variation of X is to look at how far apart X is from its mean, on average.
- One possible way to measure this variation is to consider the quantity $E[|X-\mu|]$, where $\mu = E[X]$.
- It turns out to be mathematically inconvenient to deal with this quantity.
- So a more tractable quantity is usually considered:
  The expectation of the square of the difference between X and μ.

Definition

If X is a random variable with mean μ, then the **variance** of X, denoted by $\mathrm{Var}(X)$, is defined by

$$\mathrm{Var}(X) = E[(X-\mu)^2].$$

An Alternative Formula for the Variance

- An alternative formula for $\mathrm{Var}(X)$ is derived as follows:

$$\begin{array}{rcl} \mathrm{Var}(X) & = & E[(X-\mu)^2] \\ & = & \sum_x (x-\mu)^2 p(x) \\ & = & \sum_x (x^2 - 2\mu x + \mu^2) p(x) \\ & = & \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) \\ & = & E[X^2] - 2\mu^2 + \mu^2 \\ & = & E[X^2] - \mu^2. \end{array}$$

  That is,

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2.$$

- In words, the variance of X is equal to the expected value of X^2 minus the square of its expected value.

Example

- Calculate $\text{Var}(X)$ if X represents the outcome when a fair die is rolled.

  We found in a previous example

$$\begin{aligned} E[X] &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} \\ &\quad + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= \frac{7}{2}. \end{aligned}$$

  Also,

$$\begin{aligned} E[X^2] &= 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} \\ &\quad + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} \\ &= \frac{1}{6} \cdot 91. \end{aligned}$$

  Hence, $\text{Var}(X) = \frac{91}{6} - (\frac{7}{2})^2 = \frac{35}{12}$.

Variance of Linear Function

- For any constants a and b,

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X).$$

Let $\mu = E[X]$.

From a previous corollary, $E[aX+b] = a\mu + b$.

Therefore,

$$\begin{array}{rcl} \mathrm{Var}(aX+b) & = & E[(aX+b-a\mu-b)^2] \\ & = & E[a^2(X-\mu)^2] \\ & = & a^2E[(X-\mu)^2] \\ & = & a^2\mathrm{Var}(X). \end{array}$$

Standard Deviation

(a) Returning to the terminology of mechanics:
- The means was the center of gravity of a distribution of mass;
- The variance represents the *moment of inertia*.

(b) The square root of the $\text{Var}(X)$ is called the **standard deviation** of X.

It is denoted by $\text{SD}(X)$.

That is,

$$\text{SD}(X) = \sqrt{\text{Var}(X)}.$$

Subsection 6

Bernoulli and Binomial Random Variables

Bernoulli Random Variables

- Suppose that a trial, or an experiment, whose outcome can be classified as either a **success** or a **failure** is performed.
- Let p, $0 \leq p \leq 1$ be the probability that the trial is a success.
- Consider the random variable X, such that:
  - $X = 1$ when the outcome is a success;
  - $X = 0$ when the outcome is a failure.
- The probability mass function of X is given by

$$\begin{array}{rcl} p(0) & = & P\{X = 0\} = 1 - p, \\ p(1) & = & P\{X = 1\} = p. \end{array}$$

- A random variable X is said to be a **Bernoulli random variable** if its probability mass function is given by the preceding equations for some $p \in (0, 1)$.

Binomial Random Variables

- Suppose now that n independent trials are performed.
- Each of these results in a:
  - success with probability p;
  - failure with probability $1 - p$.
- If X represents the number of successes that occur in the n trials, then X is said to be a **binomial random variable** with parameters (n, p).
- A Bernoulli random variable is just a binomial random variable with parameters $(1, p)$.
- The probability mass function of a binomial random variable having parameters (n, p) is given by

$$p(i) = \binom{n}{i} p^i (1 - p)^{n-i}, \quad i = 0, 1, \ldots, n.$$

Probability Mass of Binomial Random Variable

- Fix a particular sequence of n outcomes containing i successes and $n-i$ failures.
- The probability of this fixed sequence is, by the assumed independence of trials, $p^i(1-p)^{n-i}$.
- But the number of different sequences of n outcomes leading to i successes and $n-i$ failures is $\binom{n}{i}$.
- Hence, $p(i)=\binom{n}{i}p^i(1-p)^{n-i}$.
- By the binomial theorem, the probabilities sum to 1:

$$\sum_{i=0}^{\infty} p(i)=\sum_{i=0}^{n}\binom{n}{i}p^i(1-p)^{n-i}=[p+(1-p)]^n=1.$$

Example

- Five fair coins are flipped.

  If the outcomes are assumed independent, find the probability mass function of the number of heads obtained.

  Let X equal the number of heads (successes) that appear.

  Then X is a binomial random variable with parameters $(n = 5, p = \frac{1}{2})$.

  Hence

$$P\{X=0\} = \binom{5}{0}(\tfrac{1}{2})^0(\tfrac{1}{2})^5 = \tfrac{1}{32}, \quad P\{X=1\} = \binom{5}{1}(\tfrac{1}{2})^1(\tfrac{1}{2})^4 = \tfrac{5}{32},$$
$$P\{X=2\} = \binom{5}{2}(\tfrac{1}{2})^2(\tfrac{1}{2})^3 = \tfrac{10}{32}, \quad P\{X=3\} = \binom{5}{3}(\tfrac{1}{2})^3(\tfrac{1}{2})^2 = \tfrac{10}{32},$$
$$P\{X=4\} = \binom{5}{4}(\tfrac{1}{2})^4(\tfrac{1}{2})^1 = \tfrac{5}{32}, \quad P\{X=5\} = \binom{5}{5}(\tfrac{1}{2})^5(\tfrac{1}{2})^0 = \tfrac{1}{32}.$$

Example

- It is known that screws produced by a certain company will be defective with probability 0.01, independently of each other.

  The company sells the screws in packages of 10.

  It offers a money-back guarantee that at most 1 of the 10 is defective.

  What proportion of packages sold must the company replace?

  Let X is the number of defective screws in a package.

  Then X be a binomial random variable with parameters $(10, 0.01)$.

  Hence, the probability that a package will have to be replaced is

$$\begin{aligned} & 1 - P\{X = 0\} - P\{X = 1\} \\ & = 1 - \binom{10}{0}(0.01)^0(0.99)^{10} - \binom{10}{1}(0.01)^1(0.99)^9 \approx 0.004. \end{aligned}$$

  Thus, only 0.4 percent of the packages will have to be replaced.

Example

- The following gambling game is known as the wheel of fortune:

  A player bets on one of the numbers 1 through 6.

  Three fair dice are then rolled.
  - If the number bet by the player appears i times, $i = 1, 2, 3$, then the player wins i units;
  - If the number bet by the player does not appear on any of the dice, then the player loses 1 unit.

  Is this game fair to the player?

  We assume that the dice act independently of each other.

  Then the number of times that the number bet appears is a binomial random variable with parameters $(3, \frac{1}{6})$.

Example (Cont'd)

- Let X denote the player's winnings in the game:

$$\begin{aligned} P\{X=-1\} &= \binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3 = \frac{125}{216}, \\ P\{X=1\} &= \binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2 = \frac{75}{216}, \\ P\{X=2\} &= \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1 = \frac{15}{216}, \\ P\{X=3\} &= \binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0 = \frac{1}{216}. \end{aligned}$$

We determine whether or not this is a fair game for the player by calculating $E[X]$.

From the preceding probabilities, we obtain

$$E[X] = \frac{-125+75+30+3}{216} = \frac{-17}{216}.$$

Hence, in the long run, the player will lose 17 units per every 216 games he plays.

Example

- Suppose that a particular trait of a person is classified on the basis of one pair of genes.

  Suppose that d represents a dominant gene and r a recessive gene:
  - A person with dd genes is purely dominant;
  - One with rr is purely recessive;
  - One with rd is hybrid.

  The purely dominant and the hybrid individuals look alike.

  Children receive 1 gene from each parent.

  Suppose 2 hybrid parents have a total of 4 children.

  What is the probability that 3 of the 4 children have the outward appearance of the dominant gene?

Example (Cont'd)

- Assume that each child is equally likely to inherit either of 2 genes from each parent.

  Then the probabilities that the child of 2 hybrid parents will have *dd*, *rr* and *rd* pairs of genes are, respectively, $\frac{1}{4}, \frac{1}{4}$ and $\frac{1}{2}$.

  An offspring will have the outward appearance of the dominant gene if its gene pair is either *dd* or *rd*.

  Hence, the number of such children is binomially distributed with parameters $(4, \frac{3}{4})$.

  The desired probability is, thus,

$$\binom{4}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^1 = \frac{27}{64}.$$

Example

- Consider a jury trial in which it takes 8 of the 12 jurors to convict the defendant. That is, in order for the defendant to be convicted, at least 8 of the jurors must vote him guilty.

  Assume that jurors act independently.

  Assume, also, that each makes the right decision whether or not the defendant is guilty, with probability θ.

  What is the probability that the jury renders a correct decision?

  The problem, as stated, is incapable of solution, for there is not yet enough information.

  Suppose the defendant is innocent.

  Then the probability of the jury's rendering a correct decision is:

$$\sum_{i=5}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}.$$

Example (Cont'd)

- Suppose, on the other hand, that the defendant is guilty.

  The the probability of a correct decision is

$$\sum_{i=8}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}.$$

  Let α represent the probability that the defendant is guilty.

  We condition on whether or not the defendant is guilty.

$$\begin{aligned} P(\text{Correct Decision}) &= P(\text{Correct Decision and Innocent}) \\ &\quad + P(\text{Correct Decision and Guilty}) \\ &= P(\text{Correct Decision}|\text{Innocent})P(\text{Innocent}) \\ &\quad + P(\text{Correct Decision}|\text{Guilty})P(\text{Guilty}) \\ &= \alpha \textstyle\sum_{i=8}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i} \\ &\quad + (1-\alpha) \textstyle\sum_{i=5}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}. \end{aligned}$$

Example

- A communication system consists of n components, each of which will, independently, function with probability p.

  The total system will be able to operate effectively if at least one-half of its components function.

  For what values of p is a 5-component system more likely to operate effectively than a 3-component system?

  The number of functioning components is a binomial random variable with parameters (n, p).
  - The probability that a 5-component system will be effective is $\binom{5}{3}p^3(1-p)^2 + \binom{5}{4}p^4(1-p) + p^5$.
  - The corresponding probability for a 3-component system is $\binom{3}{2}p^2(1-p) + p^3$.

  Hence, the 5-component system is better if

$$10p^3(1-p)^2 + 5p^4(1-p) + p^5 > 3p^2(1-p) + p^3.$$

Example (Cont'd)

- We find p:

$$
\begin{gathered}
10p^3(1-p)^2 + 5p^4(1-p) + p^5 > 3p^2(1-p) + p^3 \\
10p^3 - 20p^4 + 10p^5 + 5p^4 - 5p^5 + p^5 > 3p^2 - 3p^3 + p^3 \\
6p^5 - 15p^4 + 12p^3 - 3p^2 > 0 \\
3p^2(2p^3 - 5p^2 + 4p - 1) > 0 \\
3p^2(2p^3 - 2p^2 - 3p^2 + 3p + p - 1) > 0 \\
3p^2[2p^2(p-1) - 3p(p-1) + (p-1)] > 0 \\
3p^2(p-1)(2p^2 - 3p + 1) > 0 \\
3p^2(p-1)(p-1)(2p-1) > 0 \\
3p^2(p-1)^2(2p-1) > 0 \\
2p - 1 > 0 \\
p > \frac{1}{2}.
\end{gathered}
$$

Expectation of Binomial Random Variables

- Let X be a binomial random variable with parameters n and p.
- We have
$$\begin{array}{rcl}E[X^k] & = & \sum_{i=0}^n i^k \binom{n}{i} p^i (1-p)^{n-i} \\ & = & \sum_{i=1}^n i^k \binom{n}{i} p^i (1-p)^{n-i}.\end{array}$$
- Now we use the identity $i\binom{n}{i} = n\binom{n-1}{i-1}$:
$$\begin{array}{rcl}E[X^k] & = & np \sum_{i=1}^n i^{k-1} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \\ & = & np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \\ & = & npE[(Y+1)^{k-1}],\end{array}$$

  where Y is a binomial random variable with parameters $n-1$, p.

  Setting $k=1$ in the preceding equation yields $E[X] = np$.

Variance of Binomial Random Variables

- For a binomial random variable with parameters (n, p):
  - $E[X^k] = npE[(Y+1)^{k-1}]$, where Y is a binomial random variable with parameters $(n-1, p)$;
  - $E[X] = np$.
- Set $k = 2$ in the preceding equation:

$$\begin{aligned} E[X^2] &= npE[Y+1] \\ &= np[(n-1)p+1]. \end{aligned}$$

- Since $E[X] = np$, we get

$$\begin{aligned} \mathrm{Var}(X) &= E[X^2] - (E[X])^2 \\ &= np[(n-1)p+1] - (np)^2 \\ &= np(1-p). \end{aligned}$$

Monotonicity Properties of Probability Mass

Proposition

If X is a binomial random variable with parameters (n, p), where $0 < p < 1$, then as k goes from 0 to n, $P\{X = k\}$

- first increases monotonically;
- then decreases monotonically,

reaching its largest value when k is the largest integer $\leq (n+1)p$.

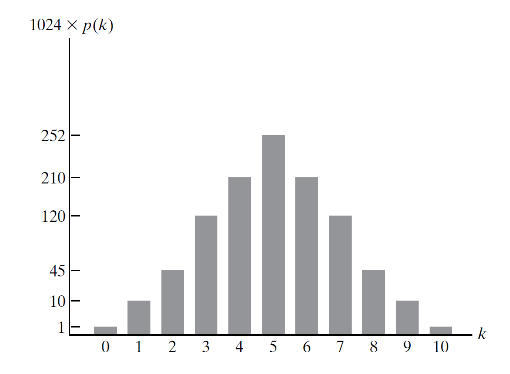
- Consider $\frac{P\{X=k\}}{P\{X=k-1\}}$ and determine for what values of k it is > 1:

$$\begin{aligned} \frac{P\{X=k\}}{P\{X=k-1\}} &= \frac{\frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}}{\frac{n!}{(n-k+1)!(k-1)!} p^{k-1} (1-p)^{n-k+1}} \\ &= \frac{(n-k+1)p}{k(1-p)}. \end{aligned}$$

Hence, $P\{X = k\} \geq P\{X = k-1\}$ iff $(n-k+1)p \geq k(1-p)$ iff $k \leq (n+1)p$.

Illustration

- As an illustration of the proposition consider the graph of the probability mass function of a binomial random variable with parameters $(10, \frac{1}{2})$.



Example

- Let X be a binomial random variable with parameters $n = 6$, $p = 0.4$.

  Recall that $P\{X = k+1\} = \frac{p}{1-p}\frac{n-k}{k+1}P\{X = k\}$.

  Starting with $P\{X = 0\} = (0.6)^6$ and working recursively, we obtain

$$\begin{aligned}
P\{X = 0\} &= (0.6)^6 \approx 0.0467, \\
P\{X = 1\} &= \tfrac{4}{6}\tfrac{6}{1}P\{X = 0\} \approx 0.1866, \\
P\{X = 2\} &= \tfrac{4}{6}\tfrac{5}{2}P\{X = 1\} \approx 0.3110, \\
P\{X = 3\} &= \tfrac{4}{6}\tfrac{4}{3}P\{X = 2\} \approx 0.2765, \\
P\{X = 4\} &= \tfrac{4}{6}\tfrac{3}{4}P\{X = 3\} \approx 0.1382, \\
P\{X = 5\} &= \tfrac{4}{6}\tfrac{2}{5}P\{X = 4\} \approx 0.0369, \\
P\{X = 6\} &= \tfrac{4}{6}\tfrac{1}{6}P\{X = 5\} \approx 0.0041.
\end{aligned}$$

Subsection 7

Poisson Random Variables

Poisson Random Variables

- A random variable X that takes on one of the values $0, 1, 2, \ldots$ is said to be a **Poisson random variable** with parameter λ if, for some $\lambda > 0$,

$$p(i) = P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!}, \quad i = 0, 1, 2, \ldots.$$

- This equation defines a probability mass function:

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1.$$

Poisson and Binomial Random Variables

- The Poisson random variable may be used as an approximation for a binomial random variable with parameters (n,p) when n is large and p is small enough so that np is of moderate size.
- Suppose that X is a binomial random variable with parameters (n,p), and let $\lambda = np$:

$$\begin{array}{rcl} P\{X=i\} & = & \frac{n!}{(n-i)!i!}p^i(1-p)^{n-i} \\ & = & \frac{n!}{(n-i)!i!}(\frac{\lambda}{n})^i(1-\frac{\lambda}{n})^{n-i} \\ & = & \frac{n(n-1)\cdots(n-i+1)}{n^i}\frac{\lambda^i}{i!}\frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^i}. \end{array}$$

- For n large and λ moderate,

$$(1-\frac{\lambda}{n})^n \approx e^{-\lambda}, \quad \frac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1, \quad (1-\frac{\lambda}{n})^i \approx 1.$$

- Hence, for n large and λ moderate, $P\{X=i\} \approx e^{-\lambda}\frac{\lambda^i}{i!}$.

Random Variables Obeying the Poisson Law

- Some examples of random variables that generally obey the Poisson probability law are as follows:
  1. The number of misprints on a page (or a group of pages) of a book;
  2. The number of people in a community who survive to age 100;
  3. The number of wrong telephone numbers that are dialed in a day;
  4. The number of packages of dog biscuits sold in a store each day;
  5. The number of customers entering a post office on a given day;
  6. The number of vacancies occurring during a year in the federal judicial system;
  7. The number of α-particles discharged in a fixed period of time from some radioactive material.

Intuitive Justification

- Each of the preceding are approximately Poisson because of the Poisson approximation to the binomial.
- Take, for instance, the number of misprints on a page of a book.

  We can suppose that there is a small probability p that each letter typed on a page will be misprinted.

  Hence, the number of misprints on a page will be approximately Poisson with $\lambda = np$, where n is the number of letters on a page.

Example

- Suppose that the number of typographical errors on a single page of this book has a Poisson distribution with parameter $\lambda = \frac{1}{2}$.

  Calculate the probability that there is at least one error on this page.

  Let X denote the number of errors on this page.

  We have

$$\begin{aligned} P\{X \geq 1\} &= 1 - P\{X = 0\} = 1 - e^{-1/2}\frac{(1/2)^0}{0!} \\ &= 1 - e^{-1/2} = \approx 0.393. \end{aligned}$$

Example

- Suppose that the probability that an item produced by a certain machine will be defective is 0.1.

  Find the probability that a sample of 10 items will contain at most 1 defective item.

  The desired probability is

$$\binom{10}{0}(0.1)^0(0.9)^{10} + \binom{10}{1}(0.1)^1(0.9)^9 = 0.7361.$$

- The Poisson approximation, with $\lambda = np = 10 \cdot 0.1 = 1$, yields the value

$$e^{-1}\frac{1^0}{0!} + e^{-1}\frac{1^1}{1!} = e^{-1} + e^{-1} \approx 0.7358.$$

Example

- An experiment consists of counting the number of α particles given off in a 1-second interval by 1 gram of radioactive material.
  If we know from past experience that, on the average, 3.2 such α particles are given off, what is a good approximation to the probability that no more than two α particles will appear?
  Think of the gram of radioactive material as consisting of a large number n of atoms.
  According to past experience, in a single second, each of these has probability $\frac{3.2}{n}$ of disintegrating and sending off an α particle.
  Thus, to a very close approximation, the number of α particles given off will be a Poisson random variable with parameter $\lambda = np = n\frac{3.2}{n} = 3.2$.
  Hence, the desired probability is
  $$P\{X \leq 2\} = e^{-3.2} + 3.2e^{-3.2} + \frac{(3.2)^2}{2}e^{-3.2} \approx 0.3799.$$

Expected Value of Poisson Random Variables

- The Poisson random variable with parameter λ approximates a binomial random variable with parameters n and p when n is large, p is small, and $\lambda = np$.
- Recall that a binomial random variable with parameters (n, p) has:
  - expected value $np = \lambda$;
  - variance $np(1-p) = \lambda(1-p) \approx \lambda$.
- Thus, it would seem that both the expected value and the variance of a Poisson random variable would equal its parameter λ.
- We now verify this result:

$$\begin{aligned} E[X] &= \sum_{i=0}^{\infty} \frac{ie^{-\lambda}\lambda^i}{i!} = \lambda \sum_{i=1}^{\infty} \frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \quad (\text{letting } j = i-1) \\ &= \lambda. \quad (\text{since } \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda}) \end{aligned}$$

Variance of Poisson Random Variables

- To determine its variance, we first compute $E[X^2]$:

$$\begin{aligned} E[X^2] &= \textstyle\sum_{i=0}^{\infty} \frac{i^2 e^{-\lambda}\lambda^i}{i!} \\ &= \lambda \textstyle\sum_{i=1}^{\infty} \frac{i e^{-\lambda}\lambda^{i-1}}{(i-1)!} \\ &= \lambda \textstyle\sum_{j=0}^{\infty} \frac{(j+1) e^{-\lambda}\lambda^{j}}{j!} \\ &= \lambda[\textstyle\sum_{j=0}^{\infty} \frac{j e^{-\lambda}\lambda^{j}}{j!} + \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^{j}}{j!}] \\ &= \lambda(\lambda+1), \end{aligned}$$

  where the final equality follows because:
  - The first sum is the expected value of a Poisson random variable with parameter λ;
  - The second is the sum of the probabilities of this random variable.

  Therefore, since we have shown that $E[X] = \lambda$, we obtain

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \lambda.$$

Weakly Dependent Events

- The Poisson distribution with parameter $\lambda = np$ is a very good approximation to the distribution of:

  the number of successes in n independent trials when each trial has probability p of being a success,

  with the proviso that:
    - the number of trials n is large;
    - the probability p of success is small.
- It remains a good approximation even when the trials are not independent, provided that *their dependence is weak*.

Example

- Recall the matching problem in which n men randomly select hats from a set consisting of one hat from each person.

  From the point of view of the number of men who select their own hat, we may regard the random selection as the result of n trials where we say that trial i is a success if person i selects his own hat, $i = 1, \ldots, n$.

  Let E_i, $i = 1, \ldots, n$, be the event $E_i = \{\text{trial } i \text{ is a success}\}$.
  - $P\{E_i\} = \frac{1}{n}$;
  - $P\{E_i | E_j\} = \frac{1}{n-1}$, $j \neq i$.

  Thus, although the events E_i, $i = 1, \ldots, n$, are not independent, their dependence, for large n, appears to be weak.

  Hence, it seems reasonable to expect that the number of successes will approximately have a Poisson distribution with $\lambda = n \times \frac{1}{n} = 1$.

  Indeed, this was verified previously.

Example

- Suppose that each of n people is equally likely to have any of the 365 days of the year as his or her birthday.

  We want to determine the probability that a set of n independent people all have different birthdays.
- A combinatorial argument was used to determine this probability, which was shown to be less than $\frac{1}{2}$ when $n = 23$.
- We approximate this probability by using the Poisson approximation:

  Consider a trial for each of the $\binom{n}{2}$ pairs of individuals i and j, $i \neq j$.

  Say trial i, j is a success if persons i and j have the same birthday.

  Let E_{ij} denote the event that trial i, j is a success.
  - The $\binom{n}{2}$ events E_{ij}, $1 \leq i < j \leq n$, are not independent;
  - However, their dependence appears to be rather weak.

Example (Cont'd)

- $P(E_{ij}) = \frac{1}{365}$.
  Hence, it is reasonable to suppose that the number of successes should approximately have a Poisson distribution with mean $\frac{\binom{n}{2}}{365} = \frac{n(n-1)}{730}$.
  Therefore,

$$\begin{array}{rcl} P\{\text{no 2 people have the same birthday}\} & = & P\{0 \text{ successes}\} \\ & \approx & \exp\left\{\frac{-n(n-1)}{730}\right\}. \end{array}$$

  To determine the smallest integer n for which this probability is less than $\frac{1}{2}$, note that $\exp\left\{\frac{-n(n-1)}{730}\right\} \leq \frac{1}{2} \Leftrightarrow \exp\left\{\frac{n(n-1)}{730}\right\} \geq 2$.
  Taking logarithms of both sides, we obtain

$$n(n-1) \geq 730 \log 2 \approx 505.997.$$

  This yields the solution $n = 23$, in agreement with the result of a previous example.

Example

- Suppose now that we wanted the probability that, among the n people, no 3 of them have their birthday on the same day.
  - This becomes a difficult combinatorial problem;
  - However, it is simple to obtain a good approximation.
- Imagine that we have a trial for each of the $\binom{n}{3}$ triplets i, j, k, where $1 \leq i < j < k \leq n$.

  Call the i, j, k trial a success if persons i, j and k all have their birthday on the same day.

  The number of successes is approximately Poisson with λ

$$\begin{aligned} \binom{n}{3} P\{i, j, k \text{ have the same birthday}\} &= \binom{n}{3}\left(\frac{1}{365}\right)^2 \\ &= \frac{n(n-1)(n-2)}{6 \times (365)^2}. \end{aligned}$$

  Hence, $P\{\text{no 3 have the same birthday}\} \approx \exp\left\{\frac{-n(n-1)(n-2)}{799350}\right\}$.

  This probability will be less than $\frac{1}{2}$ when n is such that $n(n-1)(n-2) \geq 799350 \log 2 \approx 554067.1,$ or $n \geq 84$.

Computing the Poisson Distribution Function

- If X is Poisson with parameter λ, then

$$\frac{P\{X=i+1\}}{P\{X=i\}}=\frac{e^{-\lambda}\frac{\lambda^{i+1}}{(i+1)!}}{e^{-\lambda}\frac{\lambda^i}{i!}}=\frac{\lambda}{i+1}.$$

- To compute $P\{X=i\}$, we start with $P\{X=0\}=e^{-\lambda}$.
- Then we use the preceding equation to compute successively:

$$\begin{array}{rcl} P\{X=1\} & = & \lambda P\{X=0\}; \\ P\{X=2\} & = & \frac{\lambda}{2}P\{X=1\}; \\ & \vdots & \\ P\{X=i+1\} & = & \frac{\lambda}{i+1}P\{X=i\}. \end{array}$$

Subsection 8

Expected Value of Sums of Random Variables

Sum of Random Variables

- We suppose that the sample space S is either a finite or a countably infinite set.
- For a random variable X, let $X(s)$ denote the value of X when $s \in S$ is the outcome of the experiment.
- If X and Y are both random variables, then so is their sum.

  That is, $Z = X + Y$ is also a random variable.
- Moreover, $Z(s) = X(s) + Y(s)$.

Example

- Suppose that the experiment consists of flipping a coin 5 times, with the outcome being the resulting sequence of heads and tails.

  Suppose X is the number of heads in the first 3 flips and Y is the number of heads in the final 2 flips.

  Let $Z = X + Y$.

  Then, for instance, for the outcome $s = (h, t, h, t, h)$,

$$\begin{array}{rcl} X(s) & = & 2, \\ Y(s) & = & 1, \\ Z(s) & = & X(s) + Y(s) = 3. \end{array}$$

  Clearly, the meaning is that the outcome (h, t, h, t, h) results in:
  - 2 heads in the first three flips;
  - 1 head in the final two flips;
  - A total of 3 heads in the five flips.

Probability of an Event

- Let $p(s) = P(\{s\})$ be the probability that s is the outcome of the experiment.
- Any event A can be written as the finite or countably infinite union of the mutually exclusive events $\{s\}$, $s \in A$.
- It follows by the axioms of probability that

$$P(A) = \sum_{s \in A} p(s).$$

- When $A = S$, the preceding equation gives

$$1 = \sum_{s \in S} p(s).$$

Expectation of a Random Variable

Proposition

Let X be a random variable over a sample space S. Then $E[X] = \sum_{s \in S} X(s)p(s)$.

- Suppose that the distinct values of X are x_i, $i \geq 1$.
  For each i, let S_i be the event that X is equal to x_i.
  That is, $S_i = \{s : X(s) = x_i\}$.
  Then,

$$\begin{aligned} E[X] &= \sum_i x_i P\{X = x_i\} = \sum_i x_i P(S_i) \\ &= \sum_i x_i \sum_{s \in S_i} p(s) = \sum_i \sum_{s \in S_i} x_i p(s) \\ &= \sum_i \sum_{s \in S_i} X(s)p(s) = \sum_{s \in S} X(s)p(s), \end{aligned}$$

  where the final equality follows because $S_1, S_2, \ldots$ are mutually exclusive events whose union is S.

Example

- Suppose that two independent flips of a coin that comes up heads with probability p are made.

  Let X denote the number of heads obtained:

$$\begin{array}{rcl} P(X=0) & = & P(t,t)=(1-p)^2; \\ P(X=1) & = & P(h,t)+P(t,h)=2p(1-p); \\ P(X=2) & = & P(h,h)=p^2. \end{array}$$

  Thus, by the definition of expected value,

$$E[X]=0\cdot(1-p)^2+1\cdot 2p(1-p)+2\cdot p^2=2p.$$

  This agrees with

$$\begin{array}{rcl} E[X] & = & X(h,h)p^2+X(h,t)p(1-p) \\ & & \quad +X(t,h)(1-p)p+X(t,t)(1-p)^2 \\ & = & 2p^2+p(1-p)+(1-p)p \\ & = & 2p. \end{array}$$

Expectation of a Sum of Random Variables

Corollary

For random variables $X_1, X_2, \ldots, X_n$,

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i].$$

- Let $Z = \sum_{i=1}^{n} X_i$.

  Then, by the proposition,

$$\begin{array}{rcl}
E[Z] & = & \sum_{s \in S} Z(s)p(s) \\
& = & \sum_{s \in S}(X_1(s) + X_2(s) + \cdots + X_n(s))p(s) \\
& = & \sum_{s \in S} X_1(s)p(s) + \sum_{s \in S} X_2(s)p(s) \\
& & \qquad + \cdots + \sum_{s \in S} X_n(s)p(s) \\
& = & E[X_1] + E[X_2] + \cdots + E[X_n].
\end{array}$$

Example

- Find the expected value of the sum obtained when n fair dice are rolled.

  Let X be the sum.

  We will compute $E[X]$ by using the representation $X = \sum_{i=1}^{n} X_i$, where X_i is the upturned value on die i.

  X_i is equally likely to be any of the values from 1 to 6.

  Thus,
$$E[X_i] = \sum_{i=1}^{6} i\frac{1}{6} = \frac{21}{6} = \frac{7}{2}.$$
  This yields the result
$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = 3.5n.$$

Example

- Find the expected total number of successes that result from n trials when trial i is a success with probability p_i, $i = 1, \ldots, n$.

  Let
  $$X_i = \begin{cases} 1, & \text{if trial } i \text{ is a success,} \\ 0, & \text{if trial } i \text{ is a failure.} \end{cases}$$
  Then we are interested in the expected value of $X = \sum_{i=1}^{n} X_i$.

  Note that, for all i, $E[X_i] = 1 \cdot p_i + 0 \cdot (1 - p_i) = p_i$.

  hence, we obtain
  $$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p_i.$$
- Note how this result includes as a special case the expected value of a binomial random variable: Since, it assumes independent trials and all $p_i = p$, it has mean np.

Variance of Number of Successful Trials

(a) Derive an expression for the variance of the number of successful trials in the preceding example.

(b) Then apply it to obtain the variance of a binomial random variable with parameters n and p.

(a) Let X be the number of successful trials.

Use the same representation for X - namely, $X = \sum_{i=1}^{n} X_i$:

$$\begin{array}{rcl} E[X^2] & = & E[(\sum_{i=1}^{n} X_i)(\sum_{j=1}^{n} X_j)] \\ & = & E[\sum_{i=1}^{n} X_i(X_i + \sum_{j \neq i} X_j)] \\ & = & E[\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} X_i X_j] \\ & = & \sum_{i=1}^{n} E[X_i^2] + \sum_{i=1}^{n} \sum_{j \neq i} E[X_i X_j] \\ & = & \sum_i p_i + \sum_{i=1}^{n} \sum_{j \neq i} E[X_i X_j], \end{array}$$

where the final equation used that $X_i^2 = X_i$.

Variance of Number of Successful Trials (Cont'd)

- The possible values of both X_i and X_j are 0 or 1.
  Thus,
  $$X_iX_j = \begin{cases} 1, & \text{if } X_i = 1, X_j = 1 \\ 0, & \text{otherwise.} \end{cases}$$
  Hence,
  $$E[X_iX_j] = P\{X_i = 1, X_j = 1\} = P(\text{trials } i \text{ and } j \text{ are successes}).$$

(b) If X is binomial, then, for $i \neq j$, the results of trial i and trial j are independent, with each being a success with probability p.
Therefore, $E[X_iX_j] = p^2,\ i \neq j$.
Together with the first equation, the preceding equation shows that, for a binomial random variable X, $E[X^2] = np + n(n-1)p^2$.
This implies that
$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = np + n(n-1)p^2 - n^2p^2 = np(1-p).$$

Subsection 9

Properties of the Cumulative Distribution Function

Continuity Properties of Probability

- A sequence of events $\{E_n : n \geq 1\}$ is said to be an **increasing sequence** if $E_1 \subseteq E_2 \subseteq \cdots \subseteq E_n \subseteq E_{n+1} \subseteq \cdots$.
- A sequence of events $\{E_n : n \geq 1\}$ is said to be a **decreasing sequence** if $E_1 \supseteq E_2 \supseteq \cdots \supseteq E_n \supseteq E_{n+1} \supseteq \cdots$.
- If $\{E_n : n \geq 1\}$ is an increasing sequence of events, then we define a new event, denoted by $\lim_{n\to\infty} E_n$, by $\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$.
- Similarly, if $\{E_n : n \geq 1\}$ is a decreasing sequence of events, we define $\lim_{n\to\infty} E_n$ by $\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$.

Proposition (Continuity Properties of Probability)

If $\{E_n : n \geq 1\}$ is either an increasing or a decreasing sequence of events, then

$$\lim_{n\to\infty} P(E_n) = P\left(\lim_{n\to\infty} E_n\right).$$

Properties of Cumulative Distribution Function

- Recall that, for the distribution function F of X, $F(b)$ denotes the probability that the random variable X takes on a value that is less than or equal to b.
- Following are some properties of the cumulative distribution function (c.d.f.) F:
  1. F is a nondecreasing function; that is, if $a < b$, then $F(a) \leq F(b)$.
  2. $\lim_{b \to \infty} F(b) = 1$.
  3. $\lim_{b \to -\infty} F(b) = 0$.
  4. F is right continuous. That is, for any b and any decreasing sequence b_n, $n \geq 1$, that converges to b, $\lim_{n \to \infty} F(b_n) = F(b)$.

1. Property 1: Suppose $a < b$. Then the event $\{X \leq a\}$ is contained in the event $\{X \leq b\}$. Thus, it cannot have a larger probability.
2. Property 2: Note that if b_n increases to ∞, then the events $\{X \leq b_n\}$, $n \geq 1$, are increasing events whose union is $\{X < \infty\}$. Hence, by continuity, $\lim_{n \to \infty} P\{X \leq b_n\} = P\{X < \infty\} = 1$.

Properties of Cumulative Distribution Function (Cont'd)

3. Property 3: Note that if b_n decreases to $-\infty$, then the events $\{X \leq b_n\}$, $n \geq 1$, are decreasing events whose intersection is $\{X < -\infty\} = \emptyset$.

   By continuity, $\lim_{n\to\infty} P\{X \leq b_n\} = P\{X < -\infty\} = P(\emptyset) = 0$.
4. Property 4: Note that if b_n decreases to b, then $\{X \leq b_n\}$, $n \geq 1$, are decreasing events whose intersection is $\{X \leq b\}$.

   The continuity property then, yields $\lim_n P\{X \leq b_n\} = P\{X \leq b\}$. This verifies Property 4.

Using the Cumulative Distribution Function

- All probability questions about X can be answered in terms of the c.d.f., F.
- For example,

$$P\{a < X \leq b\} = F(b) - F(a), \quad \text{for all } a < b.$$

- This equation can best be seen to hold if we write the event $\{X \leq b\}$ as the union of the mutually exclusive events $\{X \leq a\}$ and $\{a < X \leq b\}$:

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}.$$

- So

$$P\{X \leq b\} = P\{X \leq a\} + P\{a < X \leq b\}.$$

Another Useful Technique

- Suppose we want to compute the probability that X is strictly less than b.
- We can again apply the continuity property to obtain

$$\begin{aligned} P\{X < b\} &= P\left(\lim_{n\to\infty}\{X \leq b - \tfrac{1}{n}\}\right) \\ &= \lim_{n\to\infty} P(X \leq b - \tfrac{1}{n}) \\ &= \lim_{n\to\infty} F(b - \tfrac{1}{n}). \end{aligned}$$

- Note that $P\{X < b\}$ does not necessarily equal $F(b)$, since $F(b)$ also includes the probability that X equals b.

Example

- The distribution function of the random variable X is given by
$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{2}, & 0 \le x < 1 \\ \frac{2}{3}, & 1 \le x < 2 \\ \frac{11}{12}, & 2 \le x < 3 \\ 1, & 3 \le x \end{cases}.$$
Compute (a) $P\{X < 3\}$, (b) $P\{X = 1\}$, (c) $P\{X > \frac{1}{2}\}$ and (d) $P\{2 < X \le 4\}$.

(a) $P\{X < 3\} = \lim_n P\{X \le 3 - \frac{1}{n}\} = \lim_n F(3 - \frac{1}{n}) = \frac{11}{12}$.

(b) $P\{X = 1\} = P\{X \le 1\} - P\{X < 1\} = F(1) - \lim_n F(1 - \frac{1}{n}) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$.

(c) $P\{X > \frac{1}{2}\} = 1 - P\{X \le \frac{1}{2}\} = 1 - F(\frac{1}{2}) = \frac{3}{4}$.

(d) $P\{2 < X \le 4\} = F(4) - F(2) = \frac{1}{12}$.