Random Variables, Probability Distributions

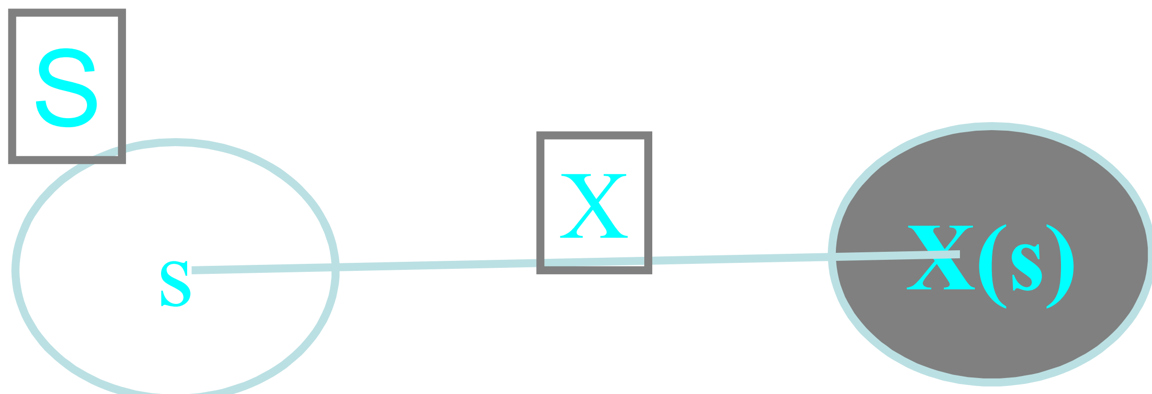
Lecture-6

Random Variables

Definition: **If S is sample space of experiment s. Then random variable X is a real valued function defined on S. In other words, for every point s in S,**

X(s) is a real number.

So, X(s)=x means that *x* is the value associated with the outcome *s* by the rv *X.*



Random Variables

Random variables are denoted by uppercase letters, such as X and Y.

Lowercase letters x and y are used to represent some particular value of the corresponding random variable.

Ex1:

When a student calls a university help desk for technical support, he/she will either immediately be able to speak to someone (*S*, for success) or will be placed on hold (*F*, for failure).

With $\mathcal{S} = \{S, F\}$, define rv X by

$$X(S) = 1 \qquad\qquad X(F) = 0$$

The rv X indicates whether (1) or not (0) the student can immediately speak to someone.

Definition

Any random variable whose only possible values are 0 and 1 is called a **Bernoulli random variable.**

Ex2:

Consider an experiment in which the number of pumps in use at each of two six-pump gas stations was determined. Define rv's X, Y, and U by

X = the total number of pumps in use at the two stations

Y = the difference between the number of pumps in use at station 1 and the number in use at station 2

U = the maximum of the numbers of pumps in use at the two stations

Ex2:

If this experiment is performed and $s = (2, 3)$ results, then

$$X((2, 3)) = 2 + 3 = 5,$$

so we say that the observed value of X is $x = 5$.

Similarly,

observed value of Y would be $y = 2 - 3 = -1$,

and observed value of U would be $u = \max(2, 3) = 3$.

Two Types of Random Variables

Two Types of Random Variables

Definition: Discrete rv

An rv whose possible values either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on ("countably" infinite).

Definition: Continuous rv

An rv is **continuous** if *both* of the following apply:

1. Its set of possible values consists either of all numbers in a single interval on the number line (possibly infinite in extent, e.g., from $-\infty$ to ∞) or all numbers in a disjoint union of such intervals (e.g., $[0, 10] \cup [20, 30]$).

2. No possible value of the variable has positive probability, that is, $P(X = c) = 0$ for any possible value c.

Ex: Discrete rv

As another example, suppose we select married couples at random and do a blood test on each person until we find a husband and wife who both have the same Rh factor.

With X = the number of blood tests to be performed, possible values of X are $D = \{2, 4, 6, 8, \ldots\}$.

Since the possible values have been listed in sequence, X is a discrete rv.

Ex: discrete rv

Suppose that we toss three coins and consider the sample space associated with the experiment.

S = {hhh, hht, hth, htt,thh, tht, tth, ttt}

X: number of tails obtained in the toss of three coins

Hence,

X(hhh)=0, X(ttt)=3,

X(tht)=2, X(htt)=2, X(tth)=2,

X(thh)=1, X(hht)=1, X(hth)=1

Ex: Continuous rv

The random variable denoting the life time of a car, when the cars's lifetime is assumed to take on any value in some interval [a, c]. So this is not discrete rv

Q:

If the sample space S is an infinite set, does this necessarily imply that any random variable X defined from will have an infinite set of possible values? If yes, say why. If not, give an example.

Ans: No. In the experiment in which a coin is tossed repeatedly until an H results, let Y = 1 if the experiment terminates with at most 5 tosses and Y = 0 otherwise. The sample space is infinite, yet Y has only two possible values.

Probability Distributions for Discrete rv

Probability Distributions for Discrete Random Variables

Probabilities assigned to various outcomes in $\mathcal{S}$ in turn determine probabilities associated with the values of any particular rv *X.*

The *probability distribution of X* says how the total probability of 1 is distributed among (allocated to) the various possible *X* values.

Suppose, for example, that a business has just purchased four laser printers, and let *X* be the number among these that require service during the warranty period.

Probability Distributions for Discrete Random Variables

Possible X values: 0, 1, 2, 3, and 4.

Probability distribution tell us how the probability of 1 is subdivided among these five possible values—how much probability is associated with X value 0, X value 1, and so on.

Use the following notation for the probabilities in the distribution:

$p(0)$ = the probability of the X value 0 = $P(X = 0)$

$p(1)$ = the probability of the X value 1 = $P(X = 1)$

and so on. In general, $p(x)$ will denote the probability assigned to the value x.

Probability Distributions for Discrete Random Variables

Definition

The **probability distribution** or **probability mass function** (pmf) of a discrete rv is defined for every number

x by $p(x) = P(X = x) = P(\text{all } s \in \mathcal{S}: X(s) = x)$.

Necessary and sufficient conditions: for p(x) to be a pmf:

1) $p(\text{x}) \geq 0$ for all x

2) $\Sigma_{\text{all possible } x}\, p(\text{x}) = 1$

Ex:

The Cal Poly Department of Statistics has a lab with six computers reserved for statistics majors.

Let X denote the number of these computers that are in use at a particular time of day.

Suppose that the probability distribution of X is as given in the following table; the first row of the table lists the possible X values and the second row gives the probability of each such value.

x	0	1	2	3	4	5	6
$p(x)$	.05	.10	.15	.25	.20	.15	.10

Ex-cont.:

Use elementary probability properties to calculate other probabilities of interest.

For example, prob. that at most 2 computers are in use is

$$P(X \leq 2) = P(X = 0 \text{ or } 1 \text{ or } 2)$$
$$= p(0) + p(1) + p(2) = .05 + .10 + .15$$

As the event *at least 3 computers are in use* is complementary to *at most 2 computers are in use,*

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - .30$$

which can, of course, also be obtained by adding together probabilities for the values, 3, 4, 5, and 6.

Cumulative Distribution Function

Cumulative Distribution Function

For some fixed value *x*, we often wish to compute the probability that the observed value of *X* will be at most *x*.

$$p(x) = \begin{cases} .500 & x = 0 \\ .167 & x = 1 \\ .333 & x = 2 \\ 0 & \text{otherwise} \end{cases}$$

The probability that *X* is at most 1 is then

$$P(X \leq 1) = p(0) + p(1) = .500 + .167 = .667$$

Cumulative Distribution Function

In this example, $X \le 1.5$ if and only if $X \le 1$, so

$$P(X \le 1.5) = P(X \le 1) = .667$$

Similarly,

$$P(X \le 0) = P(X = 0) = .5, \; P(X \le .75) = .5$$

And in fact for any x satisfying $0 \le x < 1$, $P(X \le x) = .5$.

The Cumulative Distribution Function

The largest possible X value is 2, so

$$P(X \leq 2) = 1, \quad P(X \leq 3.7) = 1, \quad P(X \leq 20.5) = 1$$

and so on.

*Note that $P(X < 1) < P(X \leq 1)$ since the latter includes the probability of the X value 1, whereas the former does not.

More generally, when X is discrete and x is a possible value of the variable, $P(X < x) < P(X \leq x)$.

The Cumulative Distribution Function

Definition

The **cumulative distribution function** (cdf) $F(x)$ of a discrete rv variable X with pmf $p(x)$ is defined for every number x by

$$F(x) = P(X \leq x) = \sum_{y:y \leq x} p(y) \qquad \textbf{(3.3)}$$

For any real number x, $F(x)$ is the probability that the observed value of X will be at most x.

The Cumulative Distribution Function

For X a discrete rv, the graph of $F(x)$ will have a jump at every possible value of X and will be flat between possible values. Such a graph is called a **step function.**

Proposition

For any two numbers a and b with $a \leq b$,

$$P(a \leq X \leq b) = F(b) - F(a-)$$

where "$a-$" represents the largest possible X value that is strictly less than a.

The Cumulative Distribution Function

In particular, if the only possible values are integers and if a and b are integers, then

$$P(a \le X \le b) = P(X = a \text{ or } a + 1 \text{ or} \ldots \text{ or } b)$$

$$= F(b) - F(a - 1)$$

Taking $a = b$ yields $P(X = a) = F(a) - F(a - 1)$ in this case.

Calculating Probabilities:

from probability density function and CDF

Interested to determine

(i) Probabilities from cumulative distribution function

(ii) Cumulative distribution functions from probability mass function & Vice versa

Ex:

A store carries flash drives with either 1 GB, 2 GB, 4 GB, 8 GB, or 16 GB of memory.

The accompanying table gives the distribution of Y = the amount of memory in a purchased drive:

y	1	2	4	8	16
$p(y)$	.05	.10	.35	.40	.10

Ex-cont.:

Let's first determine $F(y)$ for each of the five possible values of Y:

$$F(1) = P(Y \leq 1) = P(Y = 1) = .05$$

$$F(2) = P(Y \leq 2) = P(Y = 1 \text{ or } 2) = p(1) + p(2) = .15$$

$$F(4) = P(Y \leq 4) = P(Y = 1 \text{ or } 2 \text{ or } 4)$$
$$= p(1) + p(2) + p(4) = .50$$

$$F(8) = P(Y \leq 8) = p(1) + p(2) + p(4) + p(8) = .90$$

$$F(16) = P(Y \leq 16) = 1$$

Ex-cont.:

cont'd

Now for any other number y, $F(y)$ will equal the value of F at the closest possible value of Y to the left of y.

For example,

$$F(2.7) = P(Y \leq 2.7) = P(Y \leq 2)$$
$$= F(2) = .15$$

$$F(7.999) = P(Y \leq 7.999) = P(Y \leq 4)$$
$$= F(4) = .50$$

Ex-cont.:

If $y < 1$, $F(y) = 0$

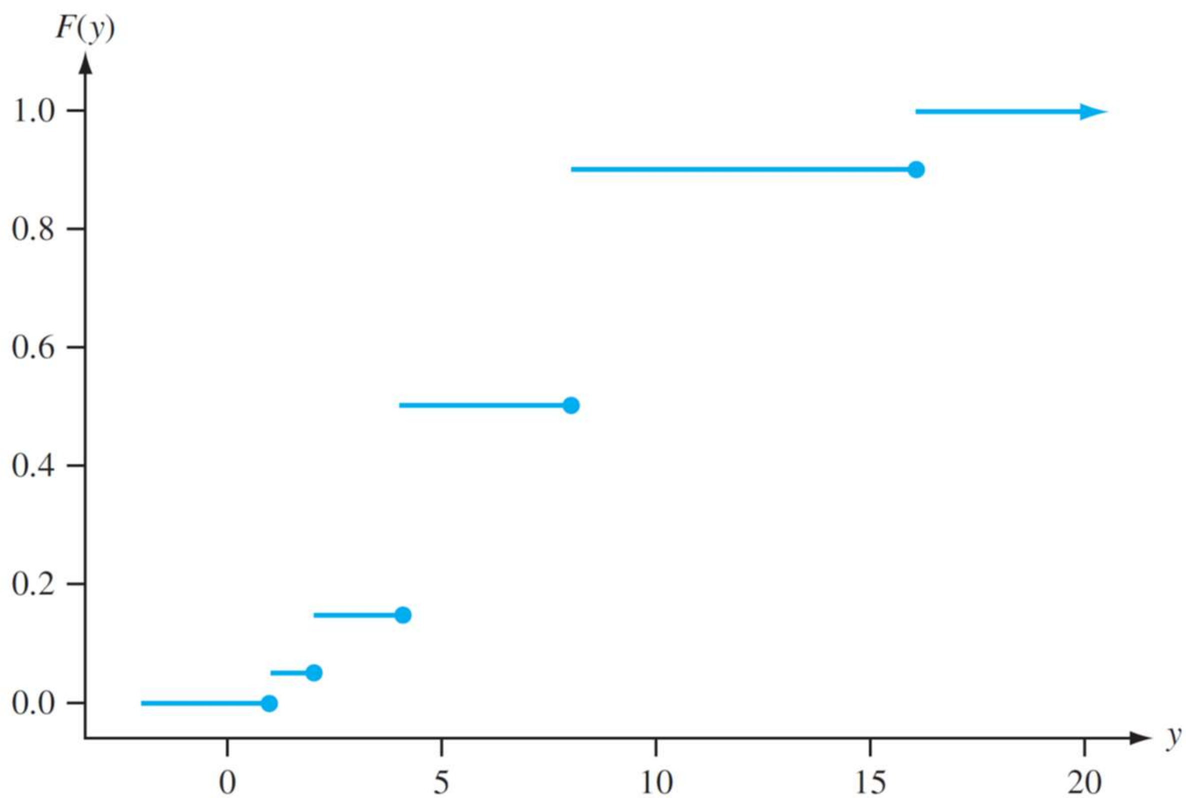
if y is at least 16, $F(y) = 1$; $F(25) = 1$.

The cdf is thus

$$F(y) = \begin{cases} 0 & y < 1 \\ .05 & 1 \le y < 2 \\ .15 & 2 \le y < 4 \\ .50 & 4 \le y < 8 \\ .90 & 8 \le y < 16 \\ 1 & 16 \le y \end{cases}$$

Ex-cont:

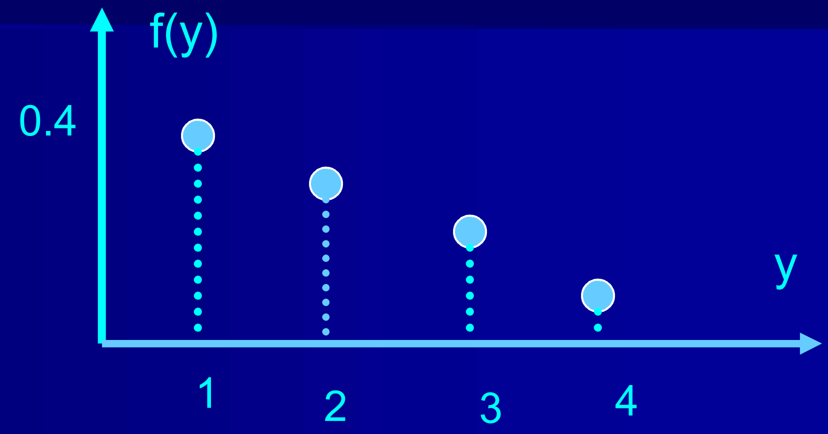
A graph of this cdf is shown below.



Ex:

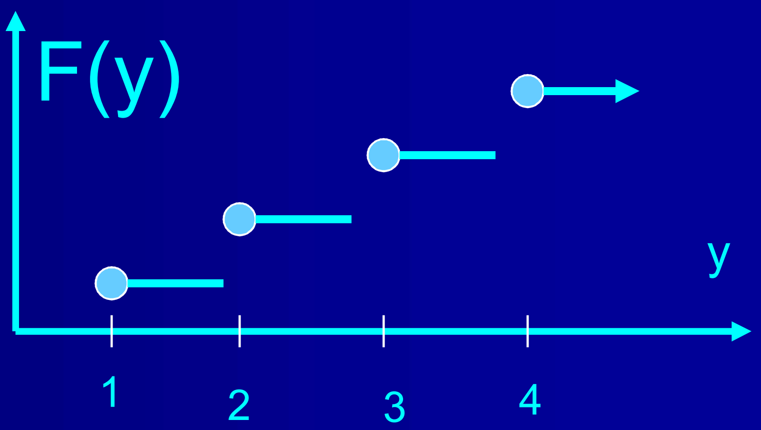
For the following probability mass function

y	1	2	3	4
f(y)	.4	.3	.2	.1



Determine the CDF

$$F(y)=\begin{cases} 0, & y<1 \\ 0.4, & 1\le y<2 \\ 0.7, & 2\le y<3 \\ 0.9 & 3\le y<4 \\ 1 & 4\le y \end{cases}$$



Ex:

Consider the following CDF

$$F(y)=\begin{cases} 0, & y<1 \\ 0.4, & 1\le y<2 \\ 0.7, & 2\le y<3 \\ 0.9 & 3\le y<4 \\ 1 & 4\le y \end{cases}$$

Y can take on 1, 2, 3, or 4.

Determine the pmf

P(Y=1)=0.4-0.0=0.4

P(Y=2)=0.7-0.4=0.3, P(Y=3)=0.9-0.7=0.2,
P(Y=4)=1.0-0.9=0.1

Hence,

y	1	2	3	4
f(y)	.4	.3	.2	.1

For a discrete random variable X

taking values $y_1, y_2, \ldots\ldots\ldots y_m$.

satisfying $y_1 < y_2 < \ldots\ldots\ldots\ldots < y_m$

If CDF F(y) is known at y_i, for $1 \le i \le m$

then $p(y_1) = F(y_1)$ &

$$p(y_i) = F(y_i) - F(y_{i-1}), 2 \le i \le m$$

Ex: Probability distribution

A contractor is required by a country planning department to submit one, two, three, four, or five forms (depending on the nature of the project) in applying for a building permit.

Let Y= Number of forms required of the next applicant.

The probability that *y* forms are required is known to be proportional to *y*—that is, $p(y)= ry$ $y=1,2,3,4,5$.

i. What is the value of *r* ?

ii. What is the probability that at most three forms are needed?

iii. What is the probability that between two and four forms(inclusive) are needed?

iv. Could $p(y)= y^2/50$, $y=1,2,3,4,5$. be the pmf of *Y*?

Expectation, Variance, Standard deviation

Lecture-7

Expected Value of X

Definition

Let X be a discrete rv with set of possible values D and pmf $p(x)$.

The **expected value** or **mean value** of X, denoted by $E(X)$ or μ_X or just μ, is

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Ex1:

Ex. Let X denotes the number of heads in a toss of two fair coins. Then find the E(X) or expected number of heads occur in tossing two fair coins.

Solution:

Ex2:

Consider a university having 15,000 students and let X = of courses for which a randomly selected student is registered. The pmf of X follows.

x	1	2	3	4	5	6	7
$p(x)$	.01	.03	.13	.25	.39	.17	.02
Number registered	150	450	1950	3750	5850	2550	300

$$\mu = 1 \cdot p(1) + 2 \cdot p(2) + \ldots + 7 \cdot p(7)$$

$$= (1)(.01) + 2(.03) + \ldots + (7)(.02)$$

$$= .01 + .06 + .39 + 1.00 + 1.95 + 1.02 + .14$$

$$= 4.57$$

Rules of expectation/Results:

If X is a random variable, then it is easy to verify the following:

(i) $E(c) = c$ (ii) $E(c\,X) = c\,E(X)$

(iii) $E(c\,X + d) = c\,E(X) + d$

(iv) $E(c\,H(X) + d\,G(X)) = c\,E(H(X)) + d\,E(G(X))$

where c, d are constants, and $H(X)$ and $G(X)$ are functions of X.

Thus, expectation respects the linearity property.

Expected value of a function

Expected value of a function

Sometimes interest will focus on expected value of some function $h(X)$ rather than on just $E(X)$.

Proposition

If the rv X has a set of possible values D and pmf $p(x)$, then the expected value of any function $h(X)$, denoted by $E[h(X)]$ or $\mu_{h(X)}$, is computed by

$$E[h(X)] = \sum_{D} h(x) \cdot p(x)$$

So, $E[h(X)]$ is computed in the same way as $E(X)$, except that $h(x)$ is substituted in place of x.

Ex3:

A computer store has purchased three computers of a certain type at \$500 apiece. It will sell them for \$1000 apiece.

The manufacturer has agreed to repurchase any computers still unsold after a specified period at \$200 apiece.

Let X denote the number of computers sold, and suppose that

$$p(0) = .1, \quad p(1) = .2, \quad p(2) = .3, \quad p(3) = .4$$

Ex3-cont.:

With $h(X)$ denoting the profit associated with selling X units, implies that

$$h(X) =$$

The expected profit is then

$$E[h(X)] =$$

Using rules of Expected Value

The $h(X)$ function of interest is quite frequently a linear function $aX + b$.

In this case, $E[h(X)]$ is easily computed from $E(X)$ as

$$E(aX + b) = a\,E(X) + b$$

(Or, using alternative notation, $\mu_{aX+b} = a\,\mu_x + b$).

Considering previous Ex 3: Since $h(X) = 800X - 900$ and $E(X) = 2$.

So, $E[h(x)] = 800\,E(X) - 900 = \$700$, as before.

Variance of X

Definition

Let X have pmf $p(x)$ and expected value μ. Then the **variance** of X, denoted by $V(X)$ or σ^2_X , or just σ^2, is

$$V(X) = \sum_D (x - \mu)^2 \cdot p(x) = E[(X - \mu)^2]$$

The **standard deviation** (SD) of X is

$$\sigma_X = \sqrt{\sigma_X^2}$$

Quantity $h(X) = (X - \mu)^2$ is the squared deviation of X from its mean, and σ^2 is the expected squared deviation.

One can say weighted average of squared deviations, where the weights are probabilities from the distribution.

Variance of X

Some observations:

If most of the probability distribution is close to μ, then σ^2 will be relatively small.

However, if there are x values far from μ that have large $p(x)$, then σ^2 will be quite large.

So, basically, σ can be interpreted as the size of a representative deviation from the mean value μ.

Ex: Let X and Y be two rv's assuming the values X = 1, 9 and Y = 4, 6. Observe that both variables have the same mean values 5. However, values of X are far away from the mean or the central value 5 in comparison to the values of Y. Thus, the mean value of a rv does not account for its variability and so variance is considered.

Ex4:

A library has an upper limit of 6 on the number of videos that can be checked out to an individual at one time.

Consider only those who check out videos, and let X denote the number of videos checked out to a randomly selected individual. The pmf of X is as follows:

x	1	2	3	4	5	6
$p(x)$	.30	.25	.15	.05	.10	.15

The expected value of X is easily seen to be $\mu = 2.85$.

Ex4-cont.:

The variance of X is then

$$V(X) = \sigma^2 = \sum_{x=1}^{6} (x - 2.85)^2 \cdot p(x)$$

$$= (1 - 2.85)^2(.30) + (2 - 2.85)^2(.25) + \ldots + (6 - 2.85)^2(.15) = 3.2275$$

The standard deviation of X is $\sigma = \sqrt{3.2275} = 1.800$.

Shortcut formula for σ^2

The number of arithmetic operations necessary to compute σ^2 can be reduced by using an alternative formula.

Proposition

$$V(X) = \sigma^2 = E(X^2) - [E(X)]^2$$

Proof:

Rules of Variance

Variance of $h(X)$ is expected value of the squared difference between $h(X)$ and its expected value:

$$V[h(X)] = \sigma^2_{h(X)} = \sum_D \{h(x) - E[h(X)]\}^2 \cdot p(x)$$

When $h(X) = aX + b$

$$h(x) - E[h(X)] = ax + b - (a\mu + b) = a(x - \mu)$$

Substituting this gives a simple relationship between

$V[h(X)]$ and $V(X)$

Rules of Variance

Proposition

$V(aX + b) = \sigma^2_{aX+b} = a^2\ \sigma^2_X$ and $\sigma_{a\,X+b} = |a| \cdot \sigma_x$

In particular,

$\sigma_{aX} = |a| \cdot \sigma_x$, $\sigma_{X+b} = \sigma_X$

The absolute value is necessary because "*a*" might be negative, yet a standard deviation cannot be.

Variance, Moment Generating Functions, Bernoulli Distribution

Lecture-8

Rules of Variance

Variance of $h(X)$ is expected value of the squared difference between $h(X)$ and its expected value:

$$V[h(X)] = \sigma^2_{h(X)} = \sum_D \{h(x) - E[h(X)]\}^2 \cdot p(x)$$

When $h(X) = a\,X + b$

$$h(x) - E[h(X)] = ax + b - (a\mu + b) = a(x - \mu)$$

Substituting this gives a simple relationship between $V[h(X)]$ and $V(X)$:

Proposition
$V(aX + b) = \sigma^2_{aX+b} = a^2\,\sigma^2_x$ and $\sigma_{a\,X + b} = |a| \cdot \sigma_x$

The absolute value is necessary because "*a*" might be negative, yet a standard deviation cannot be.

- Variance indicates the variability of a list of values.
- It is an average distance from the mean on the observations we have.
- The more different from each other our data are, the greater is the variance.

Ex. Let the rv X represent the number of automobiles that are used for social business purposes on any given workday. The probability distribution for company A is

X:	1	2	3
f(x):	0.3	0.4	0.3

and that for company B is

X:	0	1	2	3	4
f(x):	0.2	0.1	0.3	0.3	0.1

Show that the variance of the probability distribution for company B is greater than that for company A.

An appliance dealer sells three different models of upright freezers having 13.5, 15.9, and 19.1 cubic feet of storage space, respectively. Let X = the amount of storage space purchased by the next customer to buy a freezer. Suppose that X has pmf

x	13.5	15.9	19.1
$p(x)$	.2	.5	.3

a. Compute $E(X)$, $E(X^2)$, and $V(X)$.

b. If the price of a freezer having capacity X cubic feet is $25X - 8.5$, what is the expected price paid by the next customer to buy a freezer?

c. What is the variance of the price $25X - 8.5$ paid by the next customer?

d. Suppose that although the rated capacity of a freezer is X, the actual capacity is $h(X) = X - .01X^2$. What is the expected actual capacity of the freezer purchased by the next customer?

Ordinary Moments: For any positive integer r, r^{th} ordinary moments of a discrete random variable X with pmf $p(x)$ is defined as

$$E[X^r].$$

Thus, for r=1 we get mean.

Using 1^{st} and 2^{nd} ordinary moments, we can evaluate Variance.

There is a tool, ***moment generating function*** **(mgf)** which helps to evaluate all ordinary moments in one go.

Moment generating function

Definition: Let X be any random variable with density f. The mgf for X is denoted by $m_X(t)$ and is given by

$$m_X(t) = E(e^{tX})$$

provided the expectation is finite for all real numbers t in some open interval (-h, h).

It is derived from the density that allows one to calculate ordinary moments of distribution easily. It provides a unique identifier for each distribution.

Theorem: If $m_X(t)$ is the mgf for a random variable X, then

$$\left.\frac{d^r m_X(t)}{dt^r}\right|_{t=0} = E(X^r)$$

Some standard discrete distribution

Bernoulli Probability Distribution

Bernoulli Distribution

- Named after Swiss mathematician Jacob Bernoulli
- The discrete probability distribution of a random variable which takes the value 1 with probability p and the value 0 with probability $q = 1 - p$
- The probability distribution of any single experiment that asks a yes–no question; the question results in a boolean-valued outcome, a single bit whose value is success/yes/true/one with probability p and failure/no/false/zero with probability q.

Bernoulli Distribution

If X is a random variable with this distribution, then

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

The probability mass function f of this distribution, over possible outcomes x, is

$$f(x;p) = \begin{cases} 1 - p & \text{if } x = 0 \\ p & \text{if } x = 1 \\ 0 & \text{elsewhere} \end{cases}$$

This can also be expressed as

$$f(x;p) = p^x(1-p)^{1-x}; x \in \{0,1\}$$

Bernoulli Distribution

- Parameter $0 \leq p \leq 1; q = 1 - p$
- Support $k \in \{0,1\}$
- $f(x.p) = \begin{cases} 1 - p; x = 0 \\ p; x = 1 \\ 0; \text{elsewhere} \end{cases}$
- $F(x) = \begin{cases} 0; x < 0 \\ 1 - p; 0 \leq x < 1 \\ 1; x \geq 1 \end{cases}$

Bernoulli Distribution

- $E(X) = p$
- $E(X^2) = p$
- $Var(X) = pq$
- $m_X(t) = q + pe^t$