Random variables and probability distributions

Random variable (a variable, e.g. "x") that has a single numerical value, determined by chance, for each outcome of a procedure.

Probability distribution A description that gives the probability for each value of a random variable, often expressed in the format of a graph, table, or formula.

Random variables can be discrete or continuous.

discrete values like $\{1, 2, 3, \ldots\}$ countable

continuous values like interval $(1, 100)$
$1.01, 1.00125, \ldots$ uncountable

Let $P(x)$ be the probability function of a discrete random variable X. Then:

$0 \leq p(x) \leq 1$ for each value x of X.

$\sum_x P(x) = 1$ over all possible values x of X.

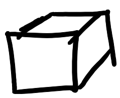
The mean, or expected value of a discrete random variable x is $\mu = E[X] = \sum x P(x)$

Think of m as the mean value of X in a very large ($\to\infty$) number of repetitions of the experiment.

Another important function associated with a random variable is the cumulative distribution function:

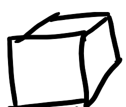
$$F(x) = p(X \leq x).$$

Ex] An experiment consists of rolling a pair of dice and letting the random variable X be the number of times 1 comes up when the dice come to rest.



dice 1

6 outcomes



dice 2

6 outcomes

36 possible outcomes

dice 1 : 2-6
dice 2 : 2-6

$\underline{X=0}$ (no ones). $p(X=0) = \frac{25}{36}$

1 to 5 for first and second dice.

$\underline{X=1}$ $\left(\begin{array}{l}\text{dice 1 has a 1}\\ \text{dice 2 } \ 2 \text{ to } 6\end{array}\right)$ or $\left(\begin{array}{l}\text{dice 1 } \ 2 \text{ to } 6\\ \text{dice 2 has 1}\end{array}\right)$

$5 + 5$

10 outcomes out of 36

$$p(X=1) = \frac{10}{36}$$

$\underline{X=2}$ $\left(\begin{array}{l}\text{dice 1 has a 1}\\ \text{dice 2 has a 1}\end{array}\right)$ $p(X=2) = \frac{1}{36}$

$P(X=0) + P(X=1) + P(X=2) = \frac{25}{36} + \frac{10}{36} + \frac{1}{36} = 1$

cumulative distribution function:

$F(x) = P(X \leq x)$

x	0	1	2
$F(x)$	$\frac{25}{36}$	$\frac{25}{36}+\frac{10}{36}$	$\frac{25}{36}+\frac{10}{36}+\frac{1}{36}$

$$\boxed{E[X] = \sum x\, p(x)} = 0\left(\frac{25}{36}\right) + 1\left(\frac{10}{36}\right) + 2\left(\frac{1}{36}\right)$$

$$= \frac{10}{36} + \frac{2}{36} = \frac{12}{36} = \frac{1}{3} = \mu$$

notice: impossible to get $\frac{1}{3}$ ones. But expected value is this. What this means is that most of the time 1 would not turn up, as we expect.

The variance of a random variable is:

$$\boxed{\sigma^2 = \sum_x (x-\mu)^2 p(x) = E[(x-\mu)^2]}$$ and $\sigma = \sqrt{\sigma^2}$ (std deviation)

Algebraically equivalent formula: last class: $\sigma^2 = \frac{\sum (x-\mu)^2}{n}$

$$\sigma^2 = \sum x_i^2 p(x_i) - \mu^2$$

$$\sigma^2 = \left(0-\frac{1}{3}\right)^2 \frac{25}{36} + \left(1-\frac{1}{3}\right)^2 \frac{10}{36} + \left(2-\frac{1}{3}\right)^2 \frac{1}{36}$$

$$\simeq 0.278 \Rightarrow \sigma \simeq 0.53$$

Some types of discrete random variables occur so often that they have been singled out and studied in detail. One of the most frequently applied distribution is the Binomial distribution.

Binomial random variables satisfy the following assumptions:

(1) there are only two possible outcomes for each trial called success and failure.

(2) the probabilities of success and failure remain the same from trial to trial.

(3) the are n trials, which are independent of each other.

A binomial random variable has a probability distribution given by:

$\tilde{p}$ prob distribution
n is # of trials, x is value

$$\tilde{p}(x) = \binom{n}{x} p^x q^{n-x} \quad \text{for } x = 0, 1, 2, \dots, n$$

$$= \underbrace{\frac{n!}{x!(n-x)!}}_{\text{binomial coefficient}} p^x q^{n-x}$$

p is prob of success

q is prob of failure

$q = 1 - p$

The binomial coefficients count the number of outcomes that give x successes.

Ex| Suppose X represent # of heads in 4 coin tosses.

X can take on five values:

$X=0$, $X=1$, $X=2$, $X=3$, or $X=4$

The probability distribution of X is binomial.

What is the probability of two heads in four tosses?

Note: multiple ways to get two heads

It makes sense for an answer to involve combinations.

Two head outcomes:

{HHTT, HTHT, HTTH, TTHH, THTH, THHT}

$$P(X=2) = \binom{4}{2} p^2 q^{4-2} \quad \text{with } p=q=0.5$$

$$= \frac{4!}{2!(4-2)!}(0.5)^2(0.5)^2 = \frac{4\cdot3\cdot2!}{2!\,2!}(0.5)^4$$

$$= 6\left(\frac{1}{2}\right)^4 = \frac{6}{16} = \frac{3}{8}$$

Notice: this is different from asking for the probability of getting two heads in two tosses.

In that case the probability is $\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$

and equivalent to $\binom{2}{2}(0.5)^2(0.5)^{2-2} = \frac{1}{4}$.

For a binomial distribution, the mean and standard deviation are: $\boxed{\mu = np}$ and $\boxed{\sigma = \sqrt{npq}}$

These can be derived from $E[X] = \sum x\tilde{p}(x)$

and $\sigma^2 = \sum (x-\mu)^2 \tilde{p}(x)$ with

$\tilde{p}(x) = \binom{n}{x} p^x q^x$

Ex) (general distribution)

x	0	1	2
$p(x)$	0.25	0.50	0.25

$\Rightarrow$

x	$p(x)$	$x\,p(x)$	$(x-\mu)^2 p(x)$	
0	0.25	0	0.25	$= (0-1)^2\, 0.25$
1	0.50	0.50	0.00	$= (1-1)^2\, 0.50$
2	0.25	0.50	0.25	$= (2-1)^2\, 0.25$
		1.00	0.50	$= \sum (x-\mu)^2 p(x) = \sigma^2$

So the mean is $\mu = \sum x p(x) = 1.00$

and $\sigma = \sqrt{\sum (x-\mu)^2 p(x)} = \sqrt{0.5}$

Ex) (binomial distribution) when the distribution is Known, the formulas specific to the distribution can be applied. Know that distribution is binomial with $p = \frac{1}{2}$.

Suppose probability distribution is as follows, with $p = \frac{1}{2}$.

X	0	1	2	3	4
$\tilde{p}(x)$	0.0625	0.25	0.375	0.25	0.0625

$m = np = 4(.5) = 2$ (using binomial formula)

$m = \sum x\tilde{p}(x) = 0(0.0625) + 1(0.25) + 2(0.375)$
$+ 3(0.25) + 4(0.0625) = 2$ (same answer, more lengthy calculation).

$\sigma^2 = npq = 4(0.5)(0.5) = 1$ (using binomial formula)

$\sigma^2 = \sum x^2\tilde{p}(x) - m^2 = 0^2(0.0625) + 1^2(0.25)$
$+ 2^2(0.375) + 3^2(0.25) + 4^2(0.0625) - 4 = 1$

notice $\sigma^2 = \sum (x-m)^2 p(x)$

Ex) Find the probability of obtaining 4 heads
(a) in 10 flips of a coin.

Let X be the random variable
X = # of heads in 10 flips of coin.
success: getting a head in a flip; failure: getting a tail $\Rightarrow$ X is binomial

$$P(X=4) = \binom{10}{4} \underset{\uparrow \atop 0.5}{p}^4 \underset{\uparrow \atop 0.5}{q}^{(10-4)} = \binom{10}{4}(0.5)^4(0.5)^6$$

$$= \binom{10}{4}(0.5)^{10}$$

$$\boxed{q = 1-p}$$

(b) Find the probability of obtaining 12 sixes in 20 throws of a die.

p is probability of getting a 6 : $p = \frac{1}{6}$

$q = 1 - \frac{1}{6} = \frac{5}{6}$

$\underline{20 = n}$ trials

X is the random variable which counts # 6's

X is binomial.

$$\underline{P(X=12)} = \binom{20}{12} p^{12} q^{20-12}$$

$$= \binom{20}{12}\left(\frac{1}{6}\right)^{12}\left(\frac{5}{6}\right)^{8}$$

(c) Find probability of at least 19 sixes in 20 tosses

$P(19) + P(20) = P(X=19) + P(X=20)$

Binomial distribution

success or failure ; $P(\text{success}) = p$; $P(\text{failure}) = q$

p and q do not change frome one trial to the next

trials are independent (called Bernoulli trials)

$$P(X=x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

<u>mean</u> $\mu = np$

<u>variance</u> $\sigma^2 = npq$

<u>std deviation</u> $\sigma = \sqrt{npq}$

Law of large numbers for Bernoulli trials:

Let X be the random variable giving the # of successes in n Bernoulli trials, so that $\frac{X}{n}$ is the proportion of successes. Then if p is the probability of success and $\varepsilon > 0$:

$$\lim_{n \to \infty} P\left(\left| \frac{X}{n} - p \right| \geq \varepsilon \right) = 0$$

In the long run, the proportion of successes $\frac{X}{n}$ will be as close as you like to the probability of success in a single trial.

Ex) Toss a fair coin three times. Find the probability that there will be
(a) 3 heads (b) 2 tails, 1 head (c) at least 1 head
(d) not more than 1 tail

In each case # trials $= n = 3$

(a) $P(\text{3 heads}) = \binom{3}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0 = \frac{1}{8}$

Define $p = \frac{1}{2}$, $q = \frac{1}{2}$.
success defined as getting a head

(b) $P(\text{2 tails and 1 head}) = \binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)$

success defined as getting a tail.

$= \frac{3!}{2!\,1!}\,\frac{1}{4\cdot 2} = \frac{3}{8}$

if opposite then, $\binom{3}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^2$ yields same result.

(c) $P(\text{at least one head}) =$

$= P(\text{1 head}) + P(\text{2 heads}) + P(\text{3 heads})$

$= 1 - P(\text{no heads})$

success defined as having a head

$= 1 - \binom{3}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3 = \frac{7}{8}$

(d) $P(\text{not more than one tail}) = P(\text{0 tails or 1 tail})$

$= P(\text{0 tails}) + P(\text{1 tail}) = \binom{3}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0 + \binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)$

defined success as getting a head.

$$= \left(\frac{1}{2}\right)^3 + \frac{3}{2}\,\frac{1}{4\cdot 2} = \frac{1}{2\cdot 2\cdot 2} + \frac{3!}{2!\,1!}\,\frac{1}{2\cdot 2\cdot 2}$$

$$= \frac{1}{8} + \frac{3}{8} = \frac{4}{8} = \frac{1}{2}$$

Ex) Suppose a fair die is tossed five times.

Find the probability that a 3 will appear

(a) twice (b) at most once (c) at least two times

$n = \#$ of trials $= 5$

(a) $P(3 \text{ occurs twice}) = P(X=2) = \binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3$

(b) $P(3 \text{ occurs at most once}) = P(X \le 1)$

$$= P(X=0) + P(X=1) = \underbrace{\binom{5}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^5}_{\text{no 3's out of 5}} + \underbrace{\binom{5}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^4}_{\text{one 3 out of 5}}$$

success : get a 3

$$= \frac{3125}{3888}$$

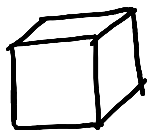
(c) $P(3 \text{ occurs at least } 2 \text{ times})$

$$= P(X \ge 2) = P(X=2) + P(X=3) + P(X=4) + P(X=5)$$

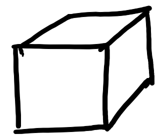
$$= 1 - P(3 \text{ occurs at most once})$$

$$= 1 - [P(X=0) + P(X=1)] = \frac{763}{3888}$$

Ex) Find the probability of getting a total of 7 at least once in three tosses of a pair of fair dice.



dice 1
1-6



dice 2
1-6

total # of outcomes $= 6 \times 6 = 36$

How to get a 7? $\begin{Bmatrix} 1 \text{ and } 6 & 3 \text{ and } 4 & 2 \text{ and } 5 \\ 6 \text{ and } 1 & 4 \text{ and } 3 & 5 \text{ and } 2 \end{Bmatrix}$

$$p(\text{sum of } 7 \text{ on two dice}) = \frac{6}{36} = \frac{1}{6}$$

X is a random variable counting the number of times a sum of 7 is attained in rolling two dice three times $\Rightarrow n = 3$ (3 trials). success: get a sum=7.

$$p(X=0) = P(\text{no } 7 \text{ in } 3 \text{ tosses}) = \binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3$$

$$= \frac{125}{216}$$

p(at least one 7 in 3 tosses) =

$$= 1 - p(X=0) = 1 - \frac{125}{216} = \frac{91}{216}.$$

Ex) If the probability of a defective bolt is 0.1, find the mean and std deviation for the # of defective bolts in a total of 400 bolts.

Let X = # of defective bolts. X is a random variable.
X follows a binomial distribution.

mean $= \mu = np = (400)(0.1) = 40$
That is, $E[X] = 40$, we expect ~40 bolts to be defective

variance $\sigma^2 = npq = (400)(0.1)(0.9) = 36$
std deviation $\sigma = \sqrt{npq} = \sqrt{36} = 6$

Poisson Random Variables

A type of probability distribution that is often useful in describing the # of events that will occur in a specific period of time or in a specific area or volume.

e.g.) the # of traffic accidents per month at a busy intersection.

characteristics of Poisson random variables:

1. The experiment consists of counting the # of times a certain event occurs during a given unit of time or in a given area or volume.

2. The probability that an event occurs in a unit

of time/area is the same for all units.

3. The # of events that occur in one unit of time/area is independent of the number that occur in other units.

The mean or expected # of events in each unit is denoted by lambda (λ).

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

$\mu = \lambda$, $\sigma^2 = \lambda$; $e = 2.71828\ldots$

The mean and variance of a Poisson random variable are both equal to λ.

Relationship to Binomial distribution:

Suppose X is Binomial rv.

Let $\lambda = np$ so that $p = \frac{\lambda}{n}$

as $n \to \infty$ can use Poisson distrib. to approx. binomial distribution!

$$P(X=x) = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)}{x!} \lambda^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

as $n \to \infty$, $P(X=x) \to \boxed{\frac{\lambda^x e^{-\lambda}}{x!} = P_{poisson}(X_{poisson} = x)}$

Ex) Ten percent of the tools produced in a certain manufacturing process turn out to be defective. Find the probability that in a sample of 10 tools chosen at random, exactly 2 will be defective, by using (a) the binomial distribution, (b) the poisson approx. to the binomial distribution.

success: tool being defective $p(\text{success}) = 0.1$

(a) $p(X=2) = \binom{10}{2}(0.1)^2(0.9)^8 \approx 0.19$

(b) $p(X=2) = \dfrac{(1)^2 e^{-1}}{2!} \approx 0.18$ ← very close

where we used $\boxed{\lambda = np = (10)(0.1) = 1}$ → b/c we use Poisson as approx to Binomial

Recap: Poisson distribution can be used to approximate the Binomial distribution.

Ex) If the probability that an individual will suffer a bad reaction from a flu shot is 0.001 determine the probability that out of 2000 individuals,

(a) exactly 3 will suffer

(b) more than 2 will suffer

$n = 2000$ is large!

Poisson prob $\approx$ binomial prob.

Suppose X is a Poisson distribution

(note, X is Binomial, but since bad reactions are rare events, we can suppose X is Poisson).

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{where } \lambda = np = (2000)(0.001) = 2$$

(a)

$$P(x=3) = \frac{2^3 e^{-2}}{3!} = 0.180$$

(much simpler calculation than for Binomial)

$$= \frac{4\cancel{8}e^{-2}}{3\cdot\cancel{2}1} = \frac{4}{3}e^{-2}$$

(b) $P(x>2) = 1-[P(x=0)+P(x=1)+P(x=2)]$

$$= 1 - \left[\frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} + \frac{2^2 e^{-2}}{2!}\right] = 1-5e^{-2}$$

≈ 0.32