Discrete Random Variables

Lecture 3

Probability & Statistics in Engineering

Dr. P.'s Clinic Consultant Module in

0909.400.01 / 0909.400.02

Clinic Consultant Module in Probability & Statistics in Engineering
Random Variables
- Discrete random variables
- Continuous random variables

Probability Distributions for Random Variables
- Probability mass (distribution) function
- Cumulative distribution function

Expected Values & Variances of Discrete Random Variables

Discrete Distribution Functions
- Binomial distribution
- Hypergeometric and negative binomial distributions
- Poisson distribution
A random variable $X$ is a function, or an association, that maps every possible event in the space $S$ of a random experiment to a real number.

- For a given sample of students, $X$=the gender of a randomly selected student $\Rightarrow$ $X$(Male)=0, $X$(Female)=1  (Bernoulli random variable)
- Quality control, The actual speed of a randomly selected chip $\Rightarrow$ $X$(chip selected) $\in [2.8 \sim 3.2]$ GHz naturally assigned continuous number
- For fishing in N. Atlantic: $X$= the next fish caught $\Rightarrow$ $X$(Salmon) = 1, $X$(Trout)=2, $X$(Sword) = 3, $X$(Seabass) = 4, $X$(Other)=5 $\leftarrow$ artificially assigned numerical values.
- For quality control: $X$=the number of chips examined before the first faulty one found $\Rightarrow$ $X$(F)=1, $X$(SF)=2, $X$(SSF)=3, $X$(SSSF)=4, ..., $X$(SSSSSSSF)=8, etc.

Random variables can be discrete, e.g., the number of heads in three consecutive coin tosses, or continuous, the weight of a class member.

Note that, random variable is just like an ordinary variable, whose value may change based on its argument, except, now, this value is random, not deterministic.
A probability mass (distribution) function is a function that tells us the probability of $x$, an observation of $X$, assuming a specific value.

Formally, the probability mass function (pmf) of a discrete random variable is defined for every possible value of $x$ as $p(x)=P(X=x)$

- In general, the random variables themselves are denoted with capital letters, and the specific values they can assume are denoted by lower case letters.
  - For the fish example, $X$ is “the next fish caught”, whereas a specific value of this variable, say for $X=\text{Salmon}$, $x=1$.

- If we have observed that of all fish we catch in N. Atlantic, 15% is salmon, 15% is trout, 10% is sword, 20% is seabass, another and 40% are all other fish, the pmf($X$) is $P(x=1)=0.15$, $P(x=2)=0.15$, $P(x=3)=0.1$, $P(x=4)=0.2$, $P(x=5)=0.4$.
The pmf also satisfies the axioms of probability, in particular

\[ P(X = x) > 0, \quad \sum_{x \in X} P(x) = 1 \]

If \( p(x) \) depends on a quantity that can be assigned differently to obtain different pmf’s, such a quantity is called the parameter of the distribution, and the collection of distributions obtained from various values of the parameter is called a family of distributions.

For Bernoulli random variable, where there are only two possible outcomes, say \( x=0 \) and \( x=1 \), the pmf can be expressed as follows, where \( \alpha \) is the parameter of family of Bernoulli distributions

\[
p(x, \alpha) = \begin{cases} \alpha & \text{if } x = 1 \\ 1 - \alpha & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}
\]
The cumulative mass (distribution) function $F(x)$ of a random variable indicates the probability of $X$ assuming a value less then or equal to $x$, i.e., $F(x)$ is the probability that the observed value of $X$ will be at most $x$.

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

$F(3) = P(X \leq 3) = P(x = 1) + P(x = 2) + P(x = 3)$

$$= p(1) + p(2) + p(3) = 0.15 + 0.15 + 0.1 = 0.4$$
Just like \textsf{cdf} can be obtained from \textsf{pmf}, so can \textsf{pmf} be obtained from \textsf{cdf}:

For any two values $a$ and $b$, $a \leq b$

\[ P(a \leq X \leq b) = F(b) - F(a^-) \]

Where $a^-$ is the largest possible value of $X$ that is strictly less then $a$. If $a$ and $b$ are integers, then

\[ P(a \leq X \leq b) = F(b) - F(a - 1) \]

Note that, if we use $a$ instead of $a^-$, we obtain

\[ P(a < X \leq b) = F(b) - F(a) \]
Let $X=$ Number of credits a Rowan student takes in a semester. We poll 1000 students and obtain the following probability mass function table.

<table>
<thead>
<tr>
<th>$x=$ # of credits</th>
<th>$\leq 12$</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>$\geq 18$</th>
</tr>
</thead>
<tbody>
<tr>
<td># of students</td>
<td>130</td>
<td>150</td>
<td>170</td>
<td>200</td>
<td>150</td>
<td>110</td>
<td>90</td>
</tr>
<tr>
<td>$P(X=x)$</td>
<td>0.13</td>
<td>0.15</td>
<td>0.17</td>
<td>0.20</td>
<td>0.15</td>
<td>0.11</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Find

a. $P(X \leq 15) = 0.65$

b. $P(13 \leq X \leq 16) = F(16) - F(14) = 0.67$

c. $F(14) = 0.45$
Expected Values of Discrete Random Variables

Let $X$ be a discrete rv with set of possible values $D$ and pmf $p(x)$. The \textit{expected value} or \textit{mean value} of $X$, is

$$E(X) = \mu_x = \sum_{x=0}^{6} x \cdot p(x)$$

The expected value of fair die is \ldots? 

Ex: At a school of 10,000 students, the number of credit cards possessed by students was as follows. The average number of credit cards held by students is then

<table>
<thead>
<tr>
<th>$x$</th>
<th># of CC</th>
<th># of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>800</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>2800</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>3800</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1600</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>600</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>300</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>
Properties of Expected Value

If the rv $X$ has the set of possible values $D$ and pmf $p(x)$, then the expected value of any function $h(x)$, is

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

If $h(x)$ is a linear function of the type $aX+b$, then

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

Therefore, for two constants $a$ and $b$

$$E(aX) = a \cdot E(X).$$

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

For any two random variables $X$ and $Y$

$$E[X + Y] = E[X] + E[Y]$$

and if $X$ and $Y$ are independent

$$E[X \cdot Y] = E[X] \cdot E[Y]$$
Let $X$ have pmf $p(x)$, and expected value $\mu$. Then the **variance** of $X$, denoted $\sigma_x^2$, or simply $\sigma^2$, is defined as the expected value of $(x-\mu)^2$.

The **standard deviation** of $X$ is the positive square root of the variance, and has the same unit as $X$.

Both the variance and the std. dev. describe the dispersion of data from the mean. If most values are close to mean, then the dispersion is small, otherwise large.
The quiz scores for a particular student are as follows: 22, 25, 20, 18, 12, 20, 24, 20, 20, 25, 24, 25, 18

We want to find the variance and standard deviation. It helps to make a frequency table so that we can obtain the pmf

<table>
<thead>
<tr>
<th>Value</th>
<th>12</th>
<th>18</th>
<th>20</th>
<th>22</th>
<th>24</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Probability</td>
<td>.08</td>
<td>.15</td>
<td>.31</td>
<td>.08</td>
<td>.15</td>
<td>.23</td>
</tr>
</tbody>
</table>

\[
\mu = \sum x \cdot p(x) = 12 \cdot 0.08 + 18 \cdot 0.15 + 20 \cdot 0.31 + 22 \cdot 0.08 + 24 \cdot 0.15 + 25 \cdot 0.23 = 21
\]

\[
\sigma^2 = E[(X - \mu)^2] = \sum_{D} (x - \mu)^2 \cdot p(x) = p_1 (x_1 - \mu)^2 + p_2 (x_2 - \mu)^2 + \ldots + p_n (x_n - \mu)^2
\]

\[
\sigma^2 = 0.08(12 - 21)^2 + 0.15(18 - 21)^2 + 0.31(20 - 21)^2 + 0.08(22 - 21)^2 + 0.15(24 - 21)^2 + 0.23(25 - 21)^2 = 13.25 \Rightarrow \sigma \approx 3.64
\]
Properties of Variance

\[ \sigma^2(aX + b) = \sigma^2_{aX+b} = a^2 \cdot \sigma^2_X \]

\[ \sigma_{aX+b} = |a| \cdot \sigma_X \]

which leads to

1. \[ \sigma^2_{aX} = a^2 \cdot \sigma^2_X, \quad \sigma_{aX} = |a| \cdot \sigma_X \]

2. \[ \sigma^2_{X+b} = \sigma^2_X \]
An experiment that satisfies the following is called a *Binomial* experiment:

- The experiment consists of a sequence of $n$ trials, where $n$ is fixed in advance of the experiment.
- The trials are identical, and each trial can result in one of the same two possible outcomes, which are denoted by success ($S$) or failure ($F$).
- The trials are independent.
- The probability of success is constant from trial to trial: denoted by $p$.

Examples: subsequent coin tosses, gender of students, chips inspected for flaw.

Note: For trials to be independent, the sample size or the number of trials must be selected from the population with replacement. Otherwise, the trials are not independent. However, if $n$ is at most 5% of the population, then sampling without replacement can still be approximated as a Binomial experiment.
There are 25 students in this class, and on average the probability of a student to get a B or better is 0.82. I will choose 10 students at random, and let X be the random variable for the number of students that obtain a score of 75 or higher.

22% of all flights in the U.S. depart delayed. Out of the 10,000 daily flights in the U.S., we are going to choose 10 flights randomly to check for delays. Let X be the number of flights that are delayed on any given day.
Binomial Distribution

- In general, we are interested in the total number of successes, rather than the outcomes of individual experiments.
- The random variable defined as
  - $X =$ the number of successes among $n$ trials of a binomial experiment is called a binomial random variable.
  - The word success is an arbitrary one; in the previous example, a delayed flight constitutes “success.” Such a trial with only two possible outcomes is called a Bernoulli trial.

- The pmf of a binomial random variable, $X$, the binomial distribution function, has two parameters
  - $n:$ the number of trials
  - $p:$ probability of success for each trial
  - and hence denoted as $b(x;n,p)$
- Example: for $n=3$, there are a possible of 8 outcomes: $SSS, SSF, SFS, SFF, FSS, FSF, FFS, FFF$ ➛ From the definition of $X$, $X(SSF)=2$, $X(FFF)=0$. 

© 2006 All Rights Reserved, Robi Polikar, Rowan University, Dept. of Electrical and Computer Engineering
For $n=4$, there are 16 outcomes. If the probability of success is $p$, and hence the probability of failure is $(1-p)$, the probability of each outcome can easily be computed: For example:

$$P(FFSF) = P(F)P(F)P(S)P(F) = (1-p)(1-p)p(1-p) = p(1-p)^3$$

<table>
<thead>
<tr>
<th>Outcome</th>
<th>$x$</th>
<th>Probability</th>
<th>Outcome</th>
<th>$x$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSSS</td>
<td>4</td>
<td>$p^4$</td>
<td>FSSS</td>
<td>3</td>
<td>$p^3(1-p)$</td>
</tr>
<tr>
<td>SSSF</td>
<td>3</td>
<td>$p^3(1-p)$</td>
<td>FSSF</td>
<td>2</td>
<td>$p^2(1-p)^2$</td>
</tr>
<tr>
<td>SSFS</td>
<td>3</td>
<td>$p^3(1-p)$</td>
<td>FSFS</td>
<td>2</td>
<td>$p^2(1-p)^2$</td>
</tr>
<tr>
<td>SSFF</td>
<td>2</td>
<td>$p^2(1-p)^2$</td>
<td>FSFF</td>
<td>1</td>
<td>$p(1-p)^3$</td>
</tr>
<tr>
<td>SFSS</td>
<td>3</td>
<td>$p^3(1-p)$</td>
<td>FFSS</td>
<td>2</td>
<td>$p^2(1-p)^2$</td>
</tr>
<tr>
<td>SFSS</td>
<td>2</td>
<td>$p^2(1-p)^2$</td>
<td>FFFS</td>
<td>1</td>
<td>$p(1-p)^3$</td>
</tr>
<tr>
<td>SFFF</td>
<td>1</td>
<td>$p(1-p)^3$</td>
<td>FFFF</td>
<td>0</td>
<td>$(1-p)^4$</td>
</tr>
</tbody>
</table>

If we want to compute say, $b(3; 4, p)$, i.e., the probability that there are three successes:

$$b(3; 4, p) = P(FSSS) + P(SFSS) + P(SSFS) + P(SSSF) = 4p^3(1-p)$$
Note that the order in which successes come is not important. We are only interested in the number of successes. Therefore, for any number of successes $x$, the binomial distribution can be written as $p^x(1-p)^{n-x}$.

Of course we need to remember that there are $\binom{n}{x}$ ways of having exactly $x$ successes out of $n$ trials. Therefore

In general, if $X$ is distributed binomially, a short hand notation is $X \sim \text{Bin}(n, p)$.

The cdf of the binomial distribution is

$$P(X \leq x) = B(x; n, p) = \sum_{y=0}^{x} b(y; n, p)$$

$x = 0, 1, 2, \ldots n$

For $X \sim \text{Bin}(n, p)$,

$$E(X) = \mu = np$$
$$\sigma^2 = np(1 - p) = npq$$

$q = 1 - p$. 
\[ x = 0:100; \]
\[ y = \text{binopdf}(x, 100, 0.5); \]
\[ \text{plot}(x, y, '+') \]

If \( p = 0.2 \), the probability of having 20 successes out of 100 trials is much higher than having, say 80 successes (which is practically zero).

However, if \( p = 0.8 \), the probability that we will have 80 successes is much higher than having, say 40 successes.
A card is drawn from a standard 52-card deck. If drawing a club is considered a success, find the probability of

a. exactly one success in 4 draws (with replacement).

\[
\binom{4}{1} \left( \frac{1}{4} \right)^1 \left( \frac{3}{4} \right)^3 \approx 0.422
\]

b. no successes in 5 draws (with replacement).

\[
\binom{5}{0} \left( \frac{1}{4} \right)^0 \left( \frac{3}{4} \right)^5 \approx 0.237
\]
If the probability of a product being free of defects is 0.82, find the probability that given 8 random samples

(a) all 8 is ok. (8 successes)

(b) All 8 is flawed (no success)

(c) at least 6 is free of defects (six successes)
Recall that binomial distribution required experiments to be independent, that is samples were selected with replacement. If the samples are selected without replacement, we obtain the *hypergeometric distribution*.

- The population or set to be sampled consists of $N$ individuals, objects, or elements (a finite population).
- Each individual can be characterized as a success ($S$) or failure ($F$), and there are $M$ successes in the population.
- A sample of $n$ individuals is selected *without replacement* in such a way that each subset of size $n$ is equally likely to be chosen.

Then the random variable of interest:

- $X = \textit{the number of successes in the sample}$ is said to have a *hypergeometric distribution*.
If $X$ is the number of $S$’s in a completely random sample of size $n$ drawn from a population consisting of $M$ $S$’s and $(N - M)$ $F$’s, then the probability distribution of $X$, is the hypergeometric distribution, given by

$$E(X) = n \cdot \frac{M}{N}, \quad \sigma^2 = \left( \frac{N - n}{N - 1} \right) \cdot n \cdot \frac{M}{N} \left( 1 - \frac{M}{N} \right)$$

with the following mean and variance

We are going to sample $n$ of the $N$ objects, and we are interested in the probability of selecting exactly $x$ successes out of the possible $M$.
Five flawed chips have accidentally passed the QC inspection and sent to a retailer in a batch of a total of 25 chips. Before selling them, the retailer decides to test the chips, by randomly picking 10 chips out of the 25 that was sent. What is the probability that a) there are two flawed chips in this sample of 10 chips; b) there are no more than two flawed chips.

This is an example of hypergeometric distribution with: \( n=10 \) sample size; \( M=5 \) number of successes (notice that a flawed chip is considered a success); \( N=25 \) total population size; \( N-M=20 \) number of failures, \( X=2 \) number of successes in the sample.

Then

For part (b),

\[
P(X \leq 2) = P(X = 0, 1, \text{or } 2) = \sum_{x=0}^{2} h(x; 10, 5, 25)
\]

\[
= 0.057 + 0.257 + 0.385 = 0.699
\]
Unlike the binomial distribution where the number of trials is fixed and the random variable is the number of successes, in the negative binomial distribution, the number of successes is fixed and the random variable is number of trials to achieve the fixed number of successes.

The negative binomial rv and distribution are based on an experiment satisfying the following four conditions:

1. The experiment consists of a sequence of independent trials.
2. Each trial can result in a success (S) or a failure (F).
3. The probability of success is constant from trial to trial, so \( P(S \text{ on trial } i) = p \) for \( i = 1, 2, 3, \ldots \)
4. The experiment continues until a total of \( r \) successes have been observed, where \( r \) is a previously specified (fixed) positive integer. In other words, the random variable \( X \) is the number of failures that precede the \( r^{th} \) success.
The pmf of the negative binomial rv \( X \) with parameters

\[ r = \text{number of } S \text{’s and } x: \text{number of failures that precede the } r^{\text{th}} \text{ success} \]

\[ p = P(S) \]

Mean and variance for nb:

\[ E(X) = \frac{r(1-p)}{p} \quad \sigma^2 = \frac{r(1-p)}{p^2} \]
The probability distribution of the number of failures before the 50th success, when the probability of success is \( p = 0.25 \). It is centered at mean \( r(1-p)/p = 50*(1-0.25)/0.25 = 150 \).
A special case of the negative binomial distribution is the geometric distribution, where \( r=1 \). Then, the random variable of interest is now

\[ X = \text{number of trials necessary to observe the first success} \]

or

\[ Y = \text{number of trials until the first success} \]

\[
\begin{align*}
nb(x;1,p) &= \binom{x+1-1}{1-1} p^1 (1-p)^x \\
\Rightarrow g(x;p) &= p (1-p)^x, \quad x = 0,1,2,\ldots \\
\text{or} \quad g(y;p) &= p (1-p)^{y-1}, \quad y = 1,2,\ldots
\end{align*}
\]

\[
\mu = \frac{1}{p} \quad \sigma^2 = \frac{1-p}{p^2}
\]
The Poisson distribution is appropriate for applications that involve counting the number of times a random event occurs in a given amount of time, distance, area, etc.

- the number of Geiger counter clicks per second,
- the number of people walking into a store in an hour,
- the number of flaws per 1000 feet of video tape.
- the number of flawed chips per 1000 chips

The Poisson distribution is a one parameter discrete distribution that takes nonnegative integer values. The parameter, \( \lambda > 0 \), typically a rate per unit time, unit area, etc., is both the mean and the variance of the distribution.

Thus, as the size of the numbers in a particular sample of Poisson random numbers gets larger, so does the variability of the numbers.
As Poisson (1837) showed, the Poisson distribution is the limiting case of a binomial distribution where $n \to \infty$ and $p \to 0$ while $np \to \lambda$.

As a rule of thumb, for the approximation of binomial distribution with Poisson distribution to be valid, the number of trials in the binomial experiment should be fairly large, say $n \geq 100$, while the probability of success per trial should be fairly low, say $p \leq 0.01$ and their product should also be small, say $np \leq 20$.

A process that satisfies the following is a Poisson process with $\lambda = \alpha t$

- There exists a parameter $\alpha > 0$ such that for any short time interval of length $\Delta t$, the probability that exactly one event is received is $\alpha \Delta t + o(\Delta t)$
- The probability of more than one event during $\Delta t$ is $o(\Delta t)$ (very very small)
- The number of events during the time interval $\Delta t$ is independent of the number that occurred prior to this time interval.
The probability distribution of the number of times an event occurs (in unit period of time), if the average rate of occurrence (in the same period) is 50.
An Example

Network packets are arriving at a communication node at an average rate of 6 per second (\( \alpha = 6 \)). We wish to find the probability that in a 0.5 second interval at least one packet is received.

The number of packets in such a period of interval has a Poisson distribution with the parameter \( \lambda = \alpha t = 6(0.5) = 3 \) (notice that \( \alpha \) is represented as a rate per second).

Then, with \( X = \text{the number of packets received in a half second interval} \), the probability that we will receive at least one packet is

\[
P(X \geq 1) = 1 - P(X = 0)
\]

\[
P(X = 0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-3} \frac{3^0}{0!} = 0.04979
\]

\[
P(X \geq 1) = 1 - 0.04979 = 0.9502
\]
Matlab can compute pmf and cdf values, can generate random numbers from a number of distributions and compute their mean and variance.

The standard syntax is

- `Namepdf(x, parameter1, parameter2,...):` compute pmf values at x
- `Namecdf(x, parameter1, parameter2, ...):` compute cdf values at x
- `Namernd(parameter1, parameter2, ...):` generate random number
- `Namestat(parameter1, parameter2,...):` compute mean and variance

where `Name` is the acronym for the distribution’s name; parameter1, parameter2, etc. are the parameters of the distribution. For the distributions we saw in this lecture the `Name` and the associated parameters (in correct order) are

- `bino:` Binomial – parameters: \( n, p \)
- `nbin:` Negative binomial – parameters: \( r, p \)
- `hyge:` Hypergeometric – parameters: \( N, M, n \)
- `geo:` Geometric – parameters: \( p \)
- `poiss:` Poisson – parameters: \( \lambda \)

Also useful:
- `randtool():` GUI for generating random numbers for all supported distributions
- `disttool():` GUI for displaying pdf/cdf of all supported distributions
Homework

Problems from Chapter 3

- 114, 122, 124, 128, 142, 149(a), 156
- Bonus 152(10%), 159(15%)