Introduction to random variables

random variable is a mapping from the sample space to real numbers. notation: X, Y, Z,...
Example: Ask a student whether she/he works part time or not.
$\mathcal{S}=\{Yes, No\}$.
X=1 if Yes, X=0 if No. Any random variable whose only possible values are 0 and 1 is called a **Bernoulli random variable**.
$Y =$ number of car accidents in a week.
Flip a coin three times. Let Z=number of heads in three flips.
$\{TTT\} \rightarrow Z = 0$
$\{HTT, THT, TTH\} \rightarrow Z = 1$
$\{HHT, HTH, THH\} \rightarrow Z = 2$
$\{HHH\} \rightarrow Z = 3$.
$W =$ the weight of a randomly selected athlete.

Discrete and continuous random variable

discrete random variable: takes finite or countably infinite values.
continuous random variable: takes all values in an interval or union of intervals.

probability distribution or **probability mass function (pmf)** of a discrete rv is defined by
$p(x) = P(X = x) = P(\text{all s} \in \mathcal{S}\text{: } X(s) = x)$. Note $p(x) \geq 0$ and $\sum_x p(x) = 1$.
Example: Suppose 20% of JMU students work part time.
$X = 1$ if yes, $X = 0$ if no.
$p(0) = P(X = 0) = P(no) = 0.8$
$p(1) = P(X = 1) = P(yes) = 0.2.$

Example: Flip a fair coin 3 times.
$X =$ number of heads in 3 flips.
$p(0) = P(X = 0) = P\{TTT\} = 1/8,$
$p(1) = P(X = 1) = P\{HTT, THT, TTH\} = 3/8,$
$p(2) = P(X = 2) = P\{HHT, HTH, THH, \} = 3/8,$
$p(3) = P(X = 3) = P\{HHH\} = 1/8.$

parameter of a distribution

Example : The probability of getting a head in each flip in p, then
$p(0) = P(X = 0) = P\{TTT\} = (1 - p)^3$,
$p(1) = P(X = 1) = P\{HTT, THT, TTH\} = 3p(1 - p)^2$,
$p(2) = P(X = 2) = P\{HHT, HTH, THH, \} = 3p^2(1 - p)$,
$p(3) = P(X = 3) = P\{HHH\} = p^3$.
p is a parameter of this distribution.

Expected values

The **expected value** or the **mean** of a discrete rv X is
$E(X) = \mu_x = \sum_x xp(x)$.
Example: The pmf of a rv X is given below:

x	1	2	3	4
p(x)	0.4	0.3	0.2	0.1

$E(X) = 1 * 0.4 + 2 * 0.3 + 3 * 0.2 + 4 * 0.1 = 2.0$

Expected value of a function of $X, h(X)$:
$\mathrm{E}(h(X)) = \sum_x h(x)p(x)$.

Example continued: Let $h(x) = \frac{1}{x}$.
$\mathrm{E}(h(X)) = E(\frac{1}{X}) = 1 * 0.4 + \frac{1}{2} * 0.3 + \frac{1}{3} * 0.2 + \frac{1}{4} * 0.1 = 0.641$.
What is $E(X^2)$?

Proposition: $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$.
Proof: $\mathrm{E}(aX + b) = \sum_x (ax + b)p(x) =$
$a \sum_x xp(x) + b \sum_x p(x) = a\mathrm{E}(X) + b$.

Two special cases of the proposition: For any constants a and b,
$\mathrm{E}(aX) = a\mathrm{E}(X), \mathrm{E}(X + b) = \mathrm{E}(X) + b$.

Variance of X

Let X have pmf $p(x)$ with mean μ. The variance of X is
$V(X) = \sigma_X^2 = \mathrm{E}(X-\mu)^2 = \sum_x (x-\mu)^2 p(x)$
$= \mathrm{E}(X^2) - \mu^2 = \sum x^2 p(x) - \mu^2$.
The standard deviation of X is $\sigma_X = \sqrt{\sigma_X^2}$.
Example: continued.
$\mathrm{E}(X^2) = 1^2 * 0.4 + 2^2 * 0.3 + 3^2 * 0.2 + 4^2 * 0.1 = 5$
$V(X) = \mathrm{E}(X^2) - \mu^2 = 5 - 2^2 = 1$.
Rules of variance:
$V(h(X)) = \sum_x [h(x) - \mathrm{E}(h(x))]^2 p(x)$.
$V(aX + b) = V(aX) = a^2 V(X)$.
proof:
$V(aX) = \mathrm{E}(aX - a\mu)^2 = \mathrm{E}a^2(X-\mu)^2 = a^2\mathrm{E}(X-\mu)^2 = a^2 V(X)$.

Moments

moments: expected values of X^r or $(X-\mu)^r$, where r is an integer.
Moments (about 0): $E(X^r)$. e.g., $E(X)$: first moment
Moments about mean: $E(X-\mu)^r$.e.g.,$E(X-\mu)^2 = V(X)$.
e.g. Third moment: $E(X^3) = \sum x^3 p(x)$.

The Binomial distribution

Binomial experiment:

1. The experiment consists of n trials.
2. Each trial has two possible outcomes, success or failure.
3. The trials are independent.
4. The probability of success, p, is constant from trial to trial.

Binomial random variable X is defined to be the number of successes out of n trials. Note the possible vales X can take are 0, 1, 2, ...n.

Find binomial probabilities

The multiple problem has 4 possible answers and only one of them is correct. If a student randomly guesses the answers for three problems, what is the probability that he can get two problems right?

$n = 3, p = 0.25$.

The three ways he can get two problems right are CCI, CIC, ICC.

Each way has a probability $0.25 * 0.25 * 0.75 = 0.047$. The probability of the three ways is: $3 * 0.25 * 0.25 * 0.75 = 0.14$.

Suppose he has to answer 5 problems. What is the probability that he can get 3 problems right?

One particular way he can get 3 right is $CCCII$, its probability is $0.25^3 * 0.75^2 = 0.0088$.

There are 10 ways he can get 3 right. The probability of these ten ways is $10 * 0.25^3 * 0.75^2 = 0.09$.

The formula for binomial distribution

$P(X=x)=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}, x=0,1,2,...n.$
$\mathrm{E}(X)=\sum xp(x)=np, V(X)=np(1-p).$
It is known that the 80% of patients who receive heart transplant surgery will survive. A hospital just performed heart transplant surgeries on 6 patients.
Find the probability that
a). All of them will survive.
b). Exactly 4 of them will survive.

$p(X = 6) = 0.8^6 = 0.262.$
$p(X = 4) = \frac{6!}{4!2!} 0.8^4 0.2^2 = 0.246.$

The Poisson distribution

A random variable is said to have a **Poisson distribution** with parameter $\lambda (\lambda > 0)$ if the pmf of X is
$p(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, 3, \cdots$.
The **mean** and **variance** of a Poisson distribution:
$E(X) = \lambda, V(X) = \lambda$.
proposition: The binomial pmf approaches the Poisson pmf if we let $n \to \infty, p \to 0$ in such a way that $np \to \lambda > 0$.

Example: Let X be the number of creatures of a certain type captured in a trap during a day. Suppose X has a Poisson distribution with $\lambda = 4.5$. i.e., on average the trap will capture 4.5 creatures a day. The probability that a trap capture 5 creatures in a day is

$$P(X = 5) = \frac{e^{-4.5}(4.5)^5}{5!} = 0.1708.$$

The probability that the trap will capture 0 creatures is

$$P(X = 0) = \frac{e^{-4.5}(4.5)^0}{0!} = e^{-4.5} = 0.011.$$

The probability that the trap will capture 12 creatures is

$$P(X = 20) = \frac{e^{-4.5}(4.5)^{12}}{12!} = 0.0016.$$

1. Let X be a discrete rv with $P(X = -2) = P(X = 2) = \frac{p}{2}$, and $P(X = 0) = 1 - p$, where p is a number between 0 and 1. Find $E(X), V(X)$ and $E(log(X + 3))$.

2. An automobile safety engineer claims that 1 in 10 automobile accidents is due to driver fatigue. If this is true, what is the probability that at least 3 of 5 automobile accidents are due to driver fatigue?

St. Petersburg Paradox

Consider a game: Toss a fair coin. The initial stake is $2 and is doubled every time a head appears. When a tail appears, the game is over and the player wins all in the pot. What is the fair price to pay to enter the game?

Exercise

It is known that a typist on average makes 2 mistakes on typing 1000 words. She is about to type an article with 2500 words. What is the probability that she will make 3 mistakes? No mistakes?