Week 3

Discrete Random Variables

ST 314
Introduction to Statistics for Engineers

Oregon State University

Introduction to Random Variables

A random variable is a quantity that takes on a real number values for event in a sample space.

Each value of a random variable has a likelihood of occurring. Letters (e.g. X and Y) are often used to denote RVs.

Types of Random Variables

Discrete	Continuous
A random variable X is **discrete** if it has a countable set of values. X may include as many elements as there are whole numbers.	A random variable X is **continuous** if it takes on all value over an interval of time or space. X has an infinite number of values between any two values

Probability Mass Function

The **Probability Mass Function** (PMF) is a probability distribution for a discrete random variable. The function assigns a probability to each possible value of x where

$$p(x) = P(\underset{\substack{\uparrow \\ \text{random variable}}}{X} = x)$$

The following properties must hold:

- $0 \le p(x) \le 1$
- $\sum_{i=1}^{D} p(x_i) = 1$

number of possible outcomes for X

European Roulette Example

The game of European roulette involves spinning a wheel with 37 slots: 18 red, 18 black, and 1 green. A ball is spun onto the wheel and will eventually land in a slot, where each slot has an equal chance of capturing the ball. Gamblers can place bets on red or black. If the ball lands on their color, they double their money. If it lands on another color, they lose their money. Suppose you play and bet \$3 on a single round.

x	\$3	-\$3
$p(x)$	$\frac{18}{37}$	$\frac{19}{37}$

Cumulative Distribution Function

The **Cumulative Distribution Function** (CDF) for values of the discrete random variable X, [redacted] is the probability that the random variable will be less than or equal to some real number *x*.

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

Expectation

Let X be a discrete random variable with D possible outcomes and a PMF $p(x)$.

The **expected value** is the value we would expect X to take on. This is the average or mean value, denoted by $E(X)$ or μ_x.

$$E(x) = \mu_x = \sum_{i=1}^{D} x_i p(x_i)$$

European Roulette Example

The game of European roulette involves spinning a wheel with 37 slots: 18 red, 18 black, and 1 green. A ball is spun onto the wheel and will eventually land in a slot, where each slot has an equal chance of capturing the ball. Gamblers can place bets on red or black. If the ball lands on their color, they double their money. If it lands on another color, they lose their money. Suppose you play and bet $3 on a single round.

x	\$3	-\$3
$p(x)$	$18/37$	$19/37$

Calculate the expected winnings on a single round.

$$E(x) = 3\left(\frac{18}{37}\right) + (-3)\left(\frac{19}{37}\right) = -0.081$$

Variance

The **variance** σ_x^2 is a measurement for variability or dispersion of X.

$$Var(X) = \sigma_x^2 = \sum_{i=1}^{n} (x_i - \mu_x)^2 \, p(x_i)$$

The **standard deviation** σ_x is the size of a typical deviation away from μ_x. The standard deviation has the same units as X and μ_x.

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{Var(X)}$$

European Roulette Example

The game of European roulette involves spinning a wheel with 37 slots: 18 red, 18 black, and 1 green. A ball is spun onto the wheel and will eventually land in a slot, where each slot has an equal chance of capturing the ball. Gamblers can place bets on red or black. If the ball lands on their color, they double their money. If it lands on another color, they lose their money. Suppose you play and bet $3 on a single round.

x	$3	-$3
$p(x)$	18/37	19/37

Calculate the standard deviation of winnings on a single round.

$\mu_x = -0.081$

$var(x) = (3+0.081)^2\left(\frac{18}{37}\right) + (-3+0.081)^2\left(\frac{19}{37}\right)$

$= 8.993$

$SD(x) = \sqrt{8.993} = \$3.00$

Linear Combinations

If W is a linear combination of X such that $W = aX + b$ where a and b are constants then the expectation and variance of W is:

$$\mu_w = E(W) = E(aX + b) = E(aX) + E(b) = aE(X) + b$$

$$\sigma_w^2 = Var(W) = Var(aX + b) = Var(aX) + \cancel{Var(b)} = a^2 Var(X)$$

European Roulette Example

The game of European roulette involves spinning a wheel with 37 slots: 18 red, 18 black, and 1 green. A ball is spun onto the wheel and will eventually land in a slot, where each slot has an equal chance of capturing the ball. Gamblers can place bets on red or black. If the ball lands on their color, they double their money. If it lands on another color, they lose their money. Suppose you play and bet $3 on a single round.

x	\$3	-\$3
$p(x)$	18/37	19/37

Suppose you bet \$1 in three different rounds. What is the expected value and standard deviation of your total winnings? $Y = \frac{1}{3}X$

one round

$E(Y) = E(\frac{1}{3}X) = \frac{1}{3}E(X) = \frac{1}{3}(-0.081) = -0.027$

All 3 rounds

$E(Y) + E(Y) + E(Y) = -0.081$

European Roulette Example

The game of European roulette involves spinning a wheel with 37 slots: 18 red, 18 black, and 1 green. A ball is spun onto the wheel and will eventually land in a slot, where each slot has an equal chance of capturing the ball. Gamblers can place bets on red or black. If the ball lands on their color, they double their money. If it lands on another color, they lose their money. Suppose you play and bet \$3 on a single round.

x	\$3	-\$3
$p(x)$	$18/37$	$19/37$

Suppose you bet \$1 in three different rounds. What is the expected value and standard deviation of your total winnings?

$Y = \frac{1}{3}X \qquad Var(X) = 8.99$

$Var(Y) = Var\left(\frac{1}{3}X\right) = \frac{1}{9}(8.99) = 1$

All three rounds:

$Var(Y) + Var(Y) + Var(Y) = 3 \qquad SD(Y) = \sqrt{3} = 1.73$

Binomial Distribution

Models the number of successful outcomes out of n Bernoulli trials with a probability of success p.

- A RV that is modeled with a binomial distribution will have come from an experiment consisting of __n__ independent Bernoulli trials.
- The outcome of interest is referred to as a __success__ and the other outcome a __failure__.
- The probability of a "success" on a single trial is denoted by __p__.
- p is a constant that does not change from trial to trial.

Suppose we flip three coins, where X is the number of times we flip a heads.

The random variable X is binomially distributed!

		HTT	HHT	
		THT	HTH	
	TTT	TTH	THH	HHH
x	0	1	2	3
$p(X = x)$	1/8	3/8	3/8	1/8

Binomial Distribution

A random variable X with a **binomial distribution** has probability mass function:

$$p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n$$

A **combination** calculates the number of ways x elements can be chosen from a sample of n total elements, order does not matter and elements cannot be repeated.

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

Binomial Expectation and Variance

A random variable X with a **binomial distribution** has expected value:

$$E(X) = np$$

A random variable X with a **binomial distribution** has variance:

$$Var(X) = np(1-p)$$

Binomial Example

On a twenty question quiz, X is the number of questions answered correctly. Each question on the quiz is a multiple choice question with four possible answers. The chance a student *guesses* the correct answer is 1/4. Each question is a Bernoulli trial.

A. What are the parameter values for the binomial distribution?

B. What is probability mass function of X?

C. How many questions should the student expect to get correct?

D. What is the standard deviation of X?

E. How likely is it that the student will get exactly 10 questions correct?

F. How likely is it that the student will get less than 20% of the questions correct?

D. $Var(X) = 20(0.25)(0.75)$

$SD(X) = \sqrt{20(0.25)(0.75)}$

Binomial Example

On a twenty question quiz, X is the number of questions answered correctly. Each question on the quiz is a multiple choice question with four possible answers. The chance a student *guesses* the correct answer is 1/4. Each question is a Bernoulli trial.

A. What are the parameter values for the binomial distribution?

B. What is probability mass function of X?

C. How many questions should the student expect to get correct?

D. What is the standard deviation of X?

E. How likely is it that the student will get exactly 10 questions correct?

F. How likely is it that the student will get less than 20% of the questions correct?

$$p(x) = P(X=x) = \binom{20}{x} 0.25^{x}\, 0.75^{20-x}$$

E. $$p(10) = P(X=10) = \underbrace{\binom{20}{10}}_{\frac{20!}{10!\,10!}} 0.25^{10}\, 0.75^{10}$$

Binomial Example

On a twenty question quiz, X is the number of questions answered correctly. Each question on the quiz is a multiple choice question with four possible answers. The chance a student *guesses* the correct answer is 1/4. Each question is a Bernoulli trial.

A. What are the parameter values for the binomial distribution?

B. What is probability mass function of X?

C. How many questions should the student expect to get correct?

D. What is the standard deviation of X?

E. How likely is it that the student will get exactly 10 questions correct?

F. How likely is it that the student will get less than 20% of the questions correct?

F. 20% is 4 questions

$$P(X<4) = P(X \leq 3)$$

$$= P(X=0) + P(X=1) + P(X=2) + P(X=3)$$

$$= 0.003 + 0.021 + 0.067 + 0.134$$

Poisson Distribution

Models the number of occurrences of an event over an interval of space or time.

- A RV that is modeled with a poisson distribution is defined by the rate parameter $\underline{\quad\lambda\quad}$.

Example: Suppose X is the number of text messages received per hour by a certain individual. On average they receive 8 text messages per hour.

Poisson Distribution

A random variable X with a **Poisson distribution** has probability mass function:

$$p(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$x = 0, 1, 2, 3, \ldots$

$$P(X=x) = \frac{8^x e^{-8}}{x!}$$

The constant e is defined by

$$e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n \approx 2.7182$$

$$P(X=5) = \frac{8^5 e^{-8}}{5!}$$

Poisson Expectation and Variance

A random variable X with a **Poisson distribution** has expected value:

$$E(X) = \lambda$$

A random variable X with a **Poisson distribution** has variance:

$$Var(X) = \lambda$$

Poisson Example

Suppose X is the number of text messages received per hour by a certain individual. On average they receive 8 text messages per hour.

A. What is the parameter value for the Poisson distribution?

$\lambda = 8$

B. What is probability mass function of *X*?

$$p(x) = \frac{8^x e^{-8}}{x!}$$

C. How many texts should the individual expect to get in an hour?

$E(X) = \lambda = 8$

D. What is the standard deviation of *X*?

$SD(X) = \sqrt{\lambda} = \sqrt{8}$