Example:

About 12% of Americans are African American. Let X = the number of African Americans in a random sample of 1500 Americans.

Then X has a ______________________ distribution.

a. What is the mean and standard deviation of the number of African Americans in the sample?

b. What is the probability that the sample contains 165 or fewer African Americans?

Normal Approximation for Binomial Distribution

When X is a count having the $B(n, p)$ distribution and n is large, then

X is approximately

Rule of Thumb: The approximation is generally good provided both $np \geq 10$ and $n(1-p) \geq 10$.

So let's try part b again ...

UE 2.4 Using Joint Probability Density Functions

The relationship between two r.v. is described in their joint distribution.

For two continuous r.v. X and Y, their **joint probability density function** $f(x, y)$ has properties

1. the surface lies on or above the horizontal plane, i.e., $f(x, y) \geq 0$ for all x, y.
2. Total **volume** under the surface is equal to 1.

For an event A, the probability of A is the total **volume** under its surface above the event A.

$$P(A) = \int_A f(x, y) dx dy$$

See Figure 2.3 (UE page 20) for a typical joint probability density function.

Example: X and Y have joint p.d.f.

$$f(x,y) = \begin{cases} \text{c} & \text{if } 0 < x < 10, 0 < y < 10 \\ 0 & \text{otherwise} \end{cases}$$

a. What is c?

b. What is the probability that $0 < X < 4$ and $0 < Y < 5$?

For discrete r.v., their joint distribution is often presented in a two-way table.

Example: Insurance Policy

An insurance company sells both a homeowner's policy and an automobile policy. For each type of policy, there is a specified deductible amount. A homeowner's policy has \$100 and \$250 deductible while an automobile policy has \$0, \$100 and \$200 deductible. The distribution of 1000 customers who purchases both policies are given below.

	Automobile			
homeowner's	\$0	\$100	\$200	Total
\$100	200	150	50	
\$250	50	250	300	
Total				

Randomly select a customer. Let $X=$ deductible amount of the homeowner's policy and $Y=$ deductible amount of the automobile policy.

1. What is the probability $X = 100$ and $Y = 0$? i.e., $P(X = 100, Y = 0) =$?

2. $P(X = 100, Y = 100) =$?

3. The joint probability distribution of X and Y is

	Y			
X	0	100	200	Total
100				
250				
Total				

4. $P(X = 100) =?$ and $P(X = 250) =?$

5. $P(Y = 0) =?$, $P(Y = 100) =?$ and $P(Y = 200) =?$

6. $P(X = 100|Y = 0) =?$ and $P(X = 250|Y = 0) =?$

Formally, let $f(x, y)$ be the joint probability function of two discrete r.v. X and Y, i.e.,

$$P(X = x, Y = y) = f(x, y)$$

Marginal Probability Density Functions are

Conditional Probability Density Functions are

Independent R.V.

Let $f(x, y)$ be the joint probability density function of two (discrete or continuous) r.v. X and Y. The marginal p.d.f. are $f(x)$ and $f(y)$.

Definition: X and Y are **independent** if

In words, two r.v. are independent if the joint p.d.f. = the product of their marginal p.d.f.

Equivalently, X and Y are **independent** if

Example: Insurance Policy

7. Are X and Y independent?

More than two r.v.

Let $X_1, \ldots, X_n$ be n r.v., with joint p.d.f. $f(x_1, \ldots, x_n)$ and marginal p.d.f. $f_1(x_1), \ldots, f_n(x_n)$.

Definition: $X_1, \ldots, X_n$ are **independent** if

UE 2.5 Covariance and Correlation

Covariance and Correlation are two important measures about the relationship between two r.v.

Definition: The **covariance** of two r.v. X and Y is

For two discrete r.v. X and Y with joint p.d.f. $f(x, y)$

Example: Insurance Policy

8. What is the covariance between X and Y?

Definition: The **correlation** of two r.v. X and Y is

Example: Insurance Policy

9. What is the correlation between X and Y?

Comments on Covariance and Correlation

- If X and Y are independent, then the covariance and correlation between them are _________
- The converse is _________ true.
- The correlation ρ must lie between ________________
- The correlation measures the degrees of ________________ association.
- The sign of correlation indicates the direction of the association.
- $\rho > 0$ indicates ________________ association and $\rho < 0$ indicates ________________ association.
- ρ measures the strength of only the linear relationship. It does not measure curved relationships.
- $\rho = 0$ means that there is ________________________

Rule for means and variances:

If X and Y are r.v., and a and b are constants, then

$$E(aX + bY) =$$

$$\text{var}\,(aX + bY) =$$

If X and Y are **independent**, then

$$\text{var}\,(aX + bY) =$$

Example: X and Y are r.v. with $E(X) = 1.5$, $E(Y) = -2.5$, $\text{var}\,(X) = 2.1$, $\text{var}\,(Y) = 1.3$.

a. What is the mean of $2X - 3Y$?

b. What is the variance of $2X - 3Y$?

In general, if $X_1, \ldots, X_n$ are r.v. and $c_1, \ldots, c_n$ are constants, then

$$E(c_1 X_1 + \cdots + c_n X_n) =$$

$$\text{var}\,(c_1 X_1 + \cdots + c_n X_n) =$$

If $X_1, \ldots, X_n$ are **independent**, then

$$\text{var}\,(c_1 X_1 + \cdots + c_n X_n) =$$

In words, the expected value of any sum is the sum of the expected values, and the variance of the sum of **independent** r.v. is the sum of their variance.

Example: Time to complete 2 chemical reactions:

Reaction 1: mean = 40 minutes, standard deviation = 2 minutes
Reaction 2: mean = 25 minutes, standard deviation = 1 minutes

There is a total of 5 minutes between the 2 reactions. The times for the two reactions are independent. Find mean and standard deviation for time to complete the process.

Section 6.4.4 Application to Random Samples

Definition: Random variables $X_1, \ldots, X_n$ form a **random sample** from a distribution if

1. they all have the same distribution and
2. they are independent of one another

If $X_1, \ldots, X_n$ are a random sample from a population with mean μ and standard deviation σ, and the mean and standard deviation of the $Sum = X_1 + \ldots + X_n$ are

Note: There is some difference between **random sample** and **simple random sample**. A **simple random sample** is a sample taken at random from a finite populations **without replacement**. Strictly speaking, a simple random sample is not a random sample. However, when the population size is much larger than the sample size, a simple random sample behaves like a random sample.