Lecture 4: Random Variables

COMP 411, Fall 2021

Victoria Manfredi

WESLEYAN UNIVERSITY

Acknowledgements: These slides are based primarily on material in the book, Introduction to Probability Models by Sheldon M. Ross, 11th edition, 2014, as well as on slides created by Vivek Srikumar (Utah) and Dan Roth (Penn), andcourses at Stanford (http://cs229.stanford.edu/) and UC Berkeley (http://ai.berkeley.edu/)

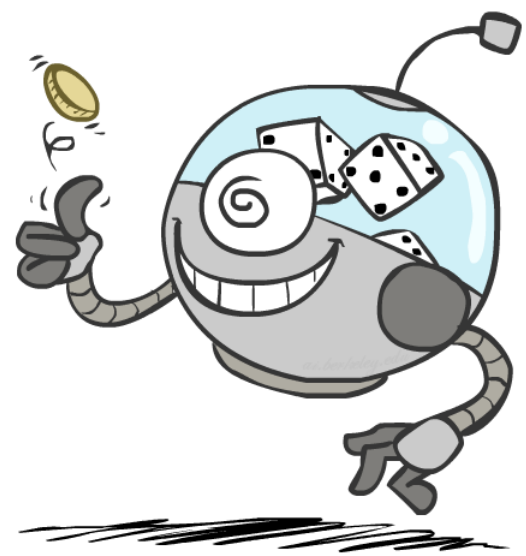
Today's Topics

Homework 2 out

- Due Thursday, September 23 by 11:59p

Random variables

- Overview
- Bayes' Rule
- Discrete: Bernoulli, Binomial
- Continuous: Uniform
- Expectation
- Joint, Independent
- Variance and Covariance



Random Variables

OVERVIEW

What is a random variable?

Suppose we perform an experiment, tossing dice

- Mainly interested in functions of outcome (e.g., is the sum of the two dice 7) rather than outcome (such as (1, 6) or (2, 5) or (3, 4) or (4, 3) or (5, 2) or (6, 1))
- These quantities of interest are known as random variables.

Random variable

- Real-valued function defined on the sample space
- We may assign probabilities to the possible values of the random variable because value of a random variable is determined by the outcome of the experiment

Example

Letting X denote the random variable that is defined as the sum of two fair dice; then

$$
\begin{aligned}
&P\{X=2\} = P\{(1,1)\} = \frac{1}{36}, \\
&P\{X=3\} = P\{(1,2),(2,1)\} = \frac{2}{36}, \\
&P\{X=4\} = P\{(1,3),(2,2),(3,1)\} = \frac{3}{36}, \\
&P\{X=5\} = P\{(1,4),(2,3),(3,2),(4,1)\} = \frac{4}{36}, \\
&P\{X=6\} = P\{(1,5),(2,4),(3,3),(4,2),(5,1)\} = \frac{5}{36}, \\
&P\{X=7\} = P\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} = \frac{6}{36}, \\
&P\{X=8\} = P\{(2,6),(3,5),(4,4),(5,3),(6,2)\} = \frac{5}{36} \\
&P\{X=9\} = P\{(3,6),(4,5),(5,4),(6,3)\} = \frac{4}{36}, \\
&P\{X=10\} = P\{(4,6),(5,5),(6,4)\} = \frac{3}{36} \\
&P\{X=11\} = P\{(5,6),(6,5)\} = \frac{2}{36}, \\
&P\{X=12\} = P\{(6,6)\} = \frac{1}{36}
\end{aligned}
$$

The random variable X can take on any integral value between 2 and 12, and the probability that it takes on each value is given above. Since X must take on one of the values 2 through 12, we must have

$$1 = P\left\{\bigcup_{i=2}^{12}\{X=n\}\right\} = \sum_{n=2}^{12} P\{X=n\}$$

Example

Suppose that our experiment consists of tossing two fair coins. Letting Y denote the number of heads appearing, then Y is a random variable taking on one of the values $0,1,2$ with respective probabilities

$$P\{Y=0\} = P\{(T,T)\} = \frac{1}{4}$$
$$P\{Y=1\} = P\{(T,H),(H,T)\} = \frac{2}{4}$$
$$P\{Y=2\} = P\{(H,H)\} = \frac{1}{4}$$

Of course, $P\{Y=0\} + P\{Y=1\} + P\{Y=2\} = 1$

Example

Suppose that we toss a coin having a probability p of coming up heads, until the first head appears. Letting N denote the number of flips required, then assuming that the outcome of successive flips are independent, N is a random variable taking on one of the values $1, 2, 3, \ldots,$, with respective probabilities

$$
\begin{aligned}
&P\{N = 1\} = P\{H\} \\
&P\{N = 2\} = P\{(T, H)\} = (1 - p)p \\
&P\{N = 3\} = P\{(T, T, H)\} = (1 - p)^2 p \\
&\vdots \\
&P\{N = n\} = P\{(T, T, \ldots, T, H)\} = (1 - p)^{n-1} p
\end{aligned}
$$

As a check, note that

$$
\begin{aligned}
P\left(\bigcup_{n=1}^{\infty} \{N = n\}\right) &= \sum_{n=1}^{\infty} P\{N = n\} \\
&= p \sum_{n=1}^{\infty} (1 - p)^{n-1} \\
&= \frac{p}{1 - (1 - p)} = 1
\end{aligned}
$$

Example

Suppose that our experiment consists of seeing how long a battery can operate before wearing down. Suppose also that we are not primarily interested in the actual lifetime of the battery but are concerned only about whether or not the battery lasts at least two years. In this case, we may define the random variable I by

$$I = \begin{cases} 1, & \text{if the lifetime of the battery is 2 or more years} \\ 0, & \text{otherwise} \end{cases}$$

If E denotes the event that the battery lasts 2 or more years, then the random variable I is known as the indicator random variable for event E. (Note that I equals 1 or 0 depending on whether or not E occurs.)

Probability Theory

BAYES FORMULA

Random variable with arity *k*

X is a random variable with arity k if it can take on exactly one value out of $x_1, x_2, \ldots, x_k$. Then

$$P(X = x_i \cap X = x_j) = 0 \quad \text{if } i \neq j$$

$$P(X = x_1 \cup X = x_2 \cup \ldots \cup X = x_k) = 1$$

$$\sum_{i=1}^{k} P(X = x_i) = 1$$

Marginalization

Let E and F be events. We may express E as $E = (E \cap F) \cup (E \cap F^c)$ because in order for a point to be in E, it must either be in both E and F, or it must be in E and not in F. Since $E \cap F$ and $E \cap F^c$ are mutually exclusive, we have that

$$\begin{aligned} P(E) &= P(E \cap F) + P(E \cap F^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)(1 - P(F)) \end{aligned}$$

The probability of the event E is a weighted average of the conditional probability of E given that F has occurred and the conditional probability of E given that F has not occurred, each conditional probability being given as much weight as the event on which it is conditioned has of occurring.

Marginalization

Can show that

$$P(Y) = P(Y \cap \{X = x_1 \cup X = x_2 \cup \ldots \cup X = x_k\})$$

$$P(Y) = \sum_{i=1}^{k} P(Y \cap X = x_i)$$

This is called marginalization over X

Bayes' Rule

$$P(F|E) = \frac{P(F \cap E)}{P(E)} = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)(1 - P(F))}$$

Can compute $P(F|E)$ from $P(E|F)$!

Why is Bayes' Rule helpful?

Lets us build one conditional from its reverse.

- Often one conditional is tricky but the other one is simple
- In the running for most important AI equation!

Example

Consider 2 urns. The first contains 2 white and 7 black balls, and the second contains 5 white and 6 black balls. We flip a fair coin and then draw a ball from the first urn or the second urn depending on whether the outcome was heads or tails. What is the conditional probability that the outcome of the toss was heads given that a white ball was selected?

Solution: Let W be the event that a white ball is drawn, and let H be the event that the coin comes up heads. Then:

$$P(H|W) = \frac{P(HW)}{P(W)} = \frac{P(W|H)P(H)}{P(W)}$$

$$= \frac{P(W|H)P(H)}{P(W|H)P(H) + P(W|H^c)P(H^c)}$$

$$= \frac{\frac{2}{9}\frac{1}{2}}{\frac{2}{9}\frac{1}{2} + \frac{5}{11}\frac{1}{2}} = \frac{22}{67}$$

Example

In answering a question on a multiple-choice test a student either knows the answer or guesses. Let p be the probability that she knows the answer and $1 - p$ the probability that she guesses. Assume that a student who guesses at the answer will be correct with probability $1/m$, where m is the number of multiple-choice alternatives. What is the conditional probability that a student knew the answer to a question given that she answered it correctly?

Solution: Let C and K denote respectively the event that the student answers the question correctly and the event that she actually knows the answer. Then:

$$P(K|C) = \frac{P(KC)}{P(C)} = \frac{P(C|K)P(K)}{P(C|K)P(K) + P(C|K^c)P(K^c)}$$

$$= \frac{p}{p + \frac{1}{m}(1-p)}$$

$$= \frac{mp}{1 + (m-1)p}$$

For example, if $m = 5$, $p = \frac{1}{2}$, then the probability that a student knew the answer to a question she correctly answered is $\frac{5}{6}$

Example

A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1 percent of the healthy persons tested. (That is, if a healthy person is tested, then, with probability 0.01, the test result will imply he has the disease.) If 0.5 percent of the population actually has the disease, what is the probability a person has the disease given that his test result is positive?

Solution: Let D be the event that the tested person has the disease, and E the event that his test result is positive. Then:

$$P(D|E) = \frac{P(DE)}{P(E)} = \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|D^c)P(D^c)}$$

$$= \frac{(0.95)(0.005)}{(0.95)(0.005) + (0.01)(0.995)}$$

$$= \frac{95}{294} \approx 0.323$$

Thus, only 32 percent of those persons whose test results are positive actually have the disease.

Bayes' rule generalized

Suppose that $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that $\bigcup_{i=1}^{n} F_i = S$. In other words, exactly one of the events $F_1, F_2, \ldots, F_n$ will occur. By writing

$$E = \bigcup_{i=1}^{n} E \cap F_i$$

and using the fact that the events $E \cap F_i,\ i = 1, \ldots, n$, are mutually exclusive, we obtain

$$\begin{aligned} P(E) &= \sum_{i=1}^{n} P(E \cap F_i) \\ &= \sum_{i=1}^{n} P(E \mid F_i) P(F_i) \end{aligned}$$

We compute $P(E)$ by first "conditioning" upon which one of the F_i occurs. Suppose now that E has occurred and we are interested in determining which one of the F_j also occurred. We then generalize Bayes' rule to:

$$P(F_j \mid E) = \frac{P(EF_j)}{P(E)} = \frac{P(E \mid F_j) P(F_j)}{\sum_{i=1}^{n} P(E \mid F_i) P(F_i)}$$

Example

You know that a certain letter is equally likely to be in any one of three different folders. Let α_i be the probability that you will find your letter upon making a quick examination of folder i if the letter is, in fact, in folder i, $i = 1,2,3$. (We may have $\alpha_i < 1$.) Suppose you look in folder 1 and do not find the letter. What is the probability that the letter is in folder 1?

Solution: Let $F_i,\ i = 1,2,3$ be the event that the letter is in folder i; and let E be the event that a search of folder 1 does not come up with the letter. We desire $P(F_1 \mid E)$. From Bayes' formula we obtain

$$P(F_1 \mid E) = \frac{P(E \mid F_1)P(F_1)}{\sum_{i=1}^{3} P(E \mid F_i)P(F_i)}$$

$$= \frac{(1-\alpha_1)\frac{1}{3}}{(1-\alpha_1)\frac{1}{3} + \frac{1}{3} + \frac{1}{3}} = \frac{1-\alpha_1}{3-\alpha_1}$$

Random Variables

DISCRETE

Discrete Random Variables

A random variable that can take on at most a countable number of possible values is said to be discrete. For a discrete random variable X, we define the probability mass function $p(a)$ of X by

$$p(a) = P\{X = a\}$$

The probability mass function $p(a)$ is positive for at most a countable number of values of a. That is, if X must assume one of the values $x_1, x_2, \ldots$, then

$$p(x_i) > 0, \qquad i = 1,2,\ldots$$

$$p(x) = 0, \qquad \text{all other values of } x$$

p(xi) > 0, i = 1,2,...

Since X must take on one of the values x_i , we have

$$\sum_{x=1}^{\infty} p(x_i) = 1$$

Discrete Random Variables

The cumulative distribution function F can be expressed in terms of $p(a)$ by

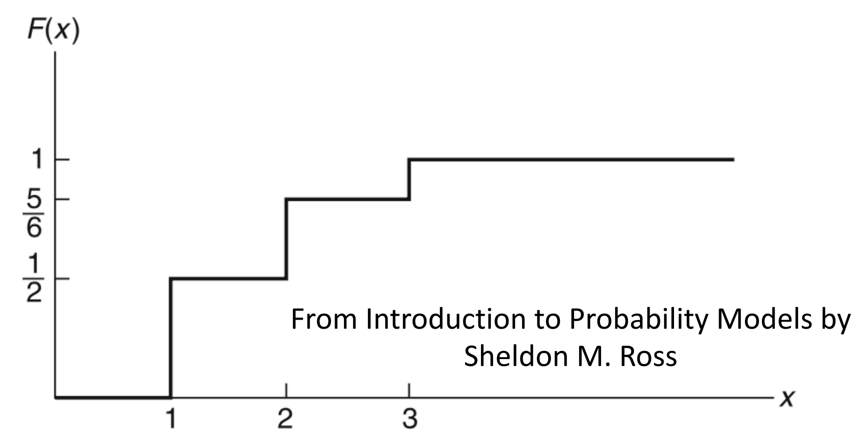
$$F(a) = \sum_{\text{all } x_i \le a} p(x_i)$$

For instance, suppose X has a probability mass function given by

$$p(1) = \frac{1}{2}, \quad p(2) = \frac{1}{3}, \quad p(3) = \frac{1}{6}$$

then, the cumulative distribution function F of X is given by

$$F(a) = \begin{cases} 0, & a < 1 \\ \frac{1}{2}, & 1 \le a < 2 \\ \frac{5}{6}, & 2 \le a < 3 \\ 1, & 3 \le a \end{cases}$$



Discrete random variables are often classified according to their probability mass functions. We now consider some of these random variables.

Bernouilli Random Variable

Suppose that a trial, or an experiment, whose outcome can be classified as either a “success” or as a “failure” is performed.

If we let X equal 1 if the outcome is a success and 0 if it is a failure, then the probability mass function of X is given by

$$p(0) = P\{X = 0\} = 1 - p$$
$$p(1) = P\{X = 1\} = p$$

where p, $0 \leq p \leq 1$ is the probability that the trial is a “success.”

A random variable X is said to be a Bernoulli random variable if its probability mass function is given by the above equation for some $p \in (0,1)$

Binomial Random Variable

Suppose that n independent trials, each of which results in a "success" with probability p and in a "failure" with probability $1 - p$, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a binomial random variable with parameters (n, p).

The probability mass function of a binomial random variable having parameters (n, p) is given by

$$p(i) = \binom{n}{i} p^i (1 - p)^{n-i}, \qquad i = 0,1,\ldots,n$$

where

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

equals the number of different groups of i objects that can be chosen from a set of n objects. The validity of this equation may be verified by first noting that the probability of any particular sequence of the n outcomes containing i successes and $n - i$ failures is, by the assumed independence of trials, $p^i(1 - p)^{n-i}$.

Random Variables

CONTINUOUS

Continuous Random Variables

Let X be such a random variable whose set of possible values is uncountable. We say that X is a continuous random variable if there exists a nonnegative function $f(x)$, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

$$P\{X \in B\} = \int_B f(x)dx$$

The function $f(x)$ is called the probability density function of the random variable X. The probability that X will be in B may be obtained by integrating the probability density function over the set B. Since X must assume some value, $f(x)$ must satisfy

$$P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)dx = 1$$

All probability statements about X can be answered in terms of $f(x)$. For instance, letting $B = [a, b]$, we obtain

$$P\{a \leq X \leq b\} = \int_a^b f(x)dx = 1$$

Continuous Random Variables

If we let $a = b$ then the probability that a continuous random variable will assume any particular value is zero.

$$P\{a \leq X \leq b\} = \int_a^a f(x)dx = 0$$

The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)dx = 0$$

Differentiating both sides of the preceding yields

$$\frac{d}{da}F(a) = f(a)$$

That is, the density is the derivative of the cumulative distribution function.

Uniform Random Variable

A random variable is said to be *uniformly distributed* over the interval (0, 1) if its probability density function is given by

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Note that the preceding is a density function since $f(x) \geq 0$ and

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{1} dx = 1$$

Since $f(x) > 0$ only when $x \in (0,1)$, it follows that X must assume a value in $(0,1)$. Also, since $f(x)$ is constant for $x \in (0,1)$, X is just as likely to be “near” any value in $(0,1)$ as any other value. To check this, note that, for any $0 < a < b < 1$,

$$P\{a \leq X \leq b\} = \int_{a}^{b} f(x)dx = b - a$$

In other words, the probability that X is in any particular subinterval of $(0,1)$ equals the length of that subinterval.

In general, we say that X is a uniform random variable on the interval (α, β) if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \text{if } \alpha < x < \beta \\ 0, & \text{otherwise} \end{cases}$$

Example

Calculate the cumulative distribution function of a random variable uniformly distributed over (α, β).

Solution: Since $F(a) = \int_{-\infty}^{a} f(x)dx$, we obtain

$$F(a) = \begin{cases} 0, & a \leq \alpha \\ \frac{a-\alpha}{\beta-\alpha}, & \alpha < a < \beta \\ 1, & a \geq \alpha \end{cases}$$

Example

If X is uniformly distributed over $(0,10)$, calculate the probability that (a) $X < 3$, (b) $X > 7$, (c) $1 < X < 6$.

Solution:

$$P\{X < 3\} = \frac{\int_0^3 dx}{10} = \frac{3}{10}$$

$$P\{X > 7\} = \frac{\int_7^{10} dx}{10} = \frac{3}{10}$$

$$P\{1 < X < 6\} = \frac{\int_1^6 dx}{10} = \frac{1}{2}$$

Random Variables

EXPECTATION

Expectation of a discrete random variable

If X is a discrete random variable having a probability mass function $p(x)$, then the expected value of X (aka, the mean) is defined by

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

In other words, the expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes that value.

Example

If the probability mass function of X is given by

$$p(1) = \frac{1}{2}, \qquad p(2) = \frac{1}{2}$$

then

$$E[X] = 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{2}\right) = \frac{3}{2}$$

is just an ordinary average of the two possible values 1 and 2 that X can assume. On the other hand, if

$$p(1) = \frac{1}{3}, \qquad p(2) = \frac{2}{3}$$

then

$$E[X] = 1\left(\frac{1}{3}\right) + 2\left(\frac{2}{3}\right) = \frac{5}{3}$$

is a weighted average of the two possible values 1 and 2 where the value 2 is given twice as much weight as the value 1 since $p(2) = 2p(1)$.

Example

Find $E[X]$ where X is the outcome when we roll a fair die.

Solution: Since $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$, we obtain

$$E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

Example

(Expectation of a Bernoulli Random Variable) Calculate $E[X]$ when X is a Bernoulli random variable with parameter p.

Solution: Since

$$p(0) = 1 - p, \quad p(1) = p,$$

then

$$E[X] = 0(1 - p) + 1(p) = p$$

Thus, the expected number of successes in a single trial is just the probability that the trial will be a success.

Expectation of a function of a random variable

Suppose we are given a random variable X and its probability distribution (that is, its probability mass function in the discrete case or its probability density function in the continuous case).

Suppose also that we are interested in calculating not the expected value of X, but the expected value of some function of X, say, $g(X)$. How do we go about doing this? One way is as follows. Since $g(X)$ is itself a random variable, it must have a probability distribution, which should be computable from a knowledge of the distribution of X. Once we have obtained the distribution of $g(X)$, we can then compute $E[g(X)]$ by the definition of the expectation.

If X is a discrete random variable with probability mass function $p(x)$, then for any real-valued function g

$$E[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$$

Example

Suppose X has the following probability mass function:

$$p(0) = 0.2,\ p(1) = 0.5,\ p(2) = 0.3$$

Calculate $E[X^2]$.

Solution: Letting $Y = X^2$, we have that Y is a random variable that can take on one of the values $0^2, 1^2, 2^2$ with respective probabilities

$$p_Y(0) = P\{Y = 0^2\} = 0.2$$

$$p_Y(1) = P\{Y = 1^2\} = 0.5$$

$$p_Y(4) = P\{Y = 2^2\} = 0.3$$

Hence,

$$E[X^2] = E[Y] = 0(0.2) + 1(0.5) + 4(0.3) = 1.7$$

Note that $1.7 = E[X^2] \neq (E[X])^2 = 1.21$

Random Variables

JOINT, INDEPENDENT

Joint cumulative probability distribution function

Thus far, we have concerned ourselves with the probability distribution of a single random variable. However, we are often interested in probability statements concerning two or more random variables. To deal with such probabilities, we define, for any two random variables X and Y, the joint cumulative probability distribution function of X and Y by

$$F(a,b) = P\{X \le a, Y \le b\}, -\infty < a,b < \infty$$

The distribution of X can be obtained from the joint distribution of X and Y as follows:

$$F_X(a) = P\{X \le a\} = P\{X \le a, Y < \infty\} = F(a,\infty)$$

Similarly, the cumulative distribution function of Y is given by

$$F_Y(b) = P\{Y \le b\} = F(\infty, b)$$

Joint probability mass function

In the case where X and Y are both discrete random variables, it is convenient to define the joint probability mass function of X and Y by $p(x, y) = P\{X = x, Y = y\}$

The probability mass function of X may be obtained from $p(x, y)$ by

$$p_X(x) = \sum_{y:p(x,y)>0} p(x, y)$$

Similarly,

$$p_Y(y) = \sum_{x:p(x,y)>0} p(x, y)$$

Expectation

If X and Y are both discrete random variables and g is a function of two variables, then

$$E[g(X,Y)] = \sum_y \sum_x g(x,y)p(x,y)$$

and for any constants a and b

$$E[aX + bY] = aE[X] + bE[Y]$$

Example

Calculate the expected sum obtained when three fair dice are rolled.

Solution: Let X denote the sum obtained. Then $X = X_1 + X_2 + X_3$ where X_i represents the value of the ith die. Thus,

$$E[X] = E[X_1] + EX_2] + E[X_3] = 3\left(\frac{7}{2}\right) = \frac{21}{2}$$

Independent random variables

The random variables X and Y are said to be independent if, for all a, b,

$$P\{X \leq a, Y \leq b\} = P\{X \leq a\}P\{Y \leq b\}$$

In other words, X and Y are independent if, for all a and b, the events $E_a = \{X \leq a\}$ and $F_b = \{Y \leq b\}$ are independent

Independent random variables

In terms of the joint distribution function F of X and Y, we have that X and Y are independent if, for all a, b,

$$F(a, b) = F_X(a)F_Y(b) \qquad \text{for all } a, b$$

When X and Y are discrete, the condition of independence reduces to

$$p(x, y) = p_X(x)p_Y(y)$$

Random Variables

VARIANCE AND COVARIANCE

More on variance

The variance of *X* is equal to the expected value of X^2 minus the square of its mean

$$Var(X) = E[X^2] - (E[X])^2 = E[X^2] - \mu_X^2$$

Proof:

$$\begin{aligned} Var(X) &= E[(X - \mu_X)^2] \\ &= E[X^2 - 2\mu_X X + \mu_X^2] \\ &= E[X^2] - 2\mu_X E[X] + \mu_X^2 \\ &= E[X^2] - \mu_X^2 \end{aligned}$$