CS 112: Computer System Modeling Fundamentals

Prof. Jenn Wortman Vaughan

April 14, 2011

Lecture 6

Reminders & Announcements

- Homework 2 has been posted on the website and is due on Tuesday, April 26
- Ethan will be traveling next week, so there will be no Monday office hours and no Friday section next week
- Tomorrow’s section will be devoted to homework 2

Last Time…

Expectation of a function: Let X be a random variable, and let Y = g(X) for an arbitrary function g. Then

$$E[Y] = E[g(X)] = \sum_{values\, k\, of\, X} g(k)\, p_X(k)$$

Last Time…

Expectation of a function: Let X be a random variable, and let Y = g(X) for an arbitrary function g. Then

$$E[Y] = E[g(X)] = \sum_{values\, k\, of\, X} g(k)\, p_X(k)$$

Linearity of expectation: Let X be a random variable, and let Y = aX + b for any scalars a and b. Then

$$E[Y] = E[aX + b] = a\,E[X] + b$$

Last Time…

The variance of a random variable measures the dispersion of X around its mean. Let $\mu = E[X]$. Then

$$\mathrm{var}(X) = E\left[(X-\mu)^2\right] = E[X^2] - \mu^2$$

Last Time…

The variance of a random variable measures the dispersion of X around its mean. Let $\mu = E[X]$. Then

$$\mathrm{var}(X) = E\left[(X-\mu)^2\right] = E[X^2] - \mu^2$$

The standard deviation of X is denoted σ_X, and is the square root of the variance. σ_X has the same units as X.

Today

What if we have multiple random variables associated with the same event?

- Joint PMFs
- Conditioning
- Independence

Today

What if we have multiple random variables associated with the same event?

- Joint PMFs
- Conditioning
- Independence

Most of this is just new notation for concepts we already know!

Joint PMFs

- Let X and Y be discrete random variables associated with the same experiment. The joint PMF of X and Y is denoted $p_{X,Y}$. For any values k of X and l of Y,

$$p_{X,Y}(k, l) = P(X=k, Y=l) = P(\{X=k\} \cap \{Y=l\})$$

Joint PMFs

- Let X and Y be discrete random variables associated with the same experiment. The joint PMF of X and Y is denoted $p_{X,Y}$. For any values *k* of X and *l* of Y,

$$p_{X,Y}(k, l) = P(X=k, Y=l) = P(\{X=k\} \cap \{Y=l\})$$

- Examples:
  - Experiment: Choose a random person in the class. X = person's height, Y = person's weight.

Joint PMFs

- Let X and Y be discrete random variables associated with the same experiment. The joint PMF of X and Y is denoted $p_{X,Y}$. For any values k of X and l of Y,

$$p_{X,Y}(k, l) = P(X=k, Y=l) = P(\{X=k\} \cap \{Y=l\})$$

- Examples:
  - Experiment: Choose a random person in the class.
    X = person's height, Y = person's weight.
  - Experiment: Run a randomized algorithm.
    X = execution time, Y = amount of memory used.

Tabular Representation

X	Y	l_1	l_2	l_3	l_4
	k_1	0.1	0.1	0	0.2
X	k_2	0.05	0.05	0.1	0
	k_3	0	0.1	0.2	0.1

Tabular Representation

Y

X	l_1	l_2	l_3	l_4
k_1	0.1	0.1	0	0.2
k_2	0.05	0.05	0.1	0
k_3	0	0.1	0.2	0.1

$p_{X,Y}(k_2, l_3) = 0.1$

Tabular Representation

X \ Y	l_1	l_2	l_3	l_4
k_1	0.1	0.1	0	0.2
k_2	0.05	0.05	0.1	0
k_3	0	0.1	0.2	0.1

$p_{X,Y}(k_2, l_3) = 0.1$

Is this a valid joint PMF? How do we know?

Marginal PMFs

We can compute the PMFs of X and Y from their joint PMF:

$$p_X(k) = \sum_l p_{X,Y}(k,l) \qquad\qquad p_Y(l) = \sum_k p_{X,Y}(k,l)$$

Marginal PMFs

We can compute the PMFs of X and Y from their joint PMF:

$$p_X(k) = \sum_l p_{X,Y}(k,l) \qquad p_Y(l) = \sum_k p_{X,Y}(k,l)$$

	l_1	l_2	l_3	l_4
k_1	0.1	0.1	0	0.2
k_2	0.05	0.05	0.1	0
k_3	0	0.1	0.2	0.1

Marginal PMFs

We can compute the PMFs of X and Y from their joint PMF:

$$p_X(k) = \sum_l p_{X,Y}(k,l) \qquad\qquad p_Y(l) = \sum_k p_{X,Y}(k,l)$$

	l_1	l_2	l_3	l_4	$p_X(k)$
k_1	0.1	0.1	0	0.2	0.4
k_2	0.05	0.05	0.1	0	0.2
k_3	0	0.1	0.2	0.1	0.4

Marginal PMFs

We can compute the PMFs of X and Y from their joint PMF:

$$p_X(k) = \sum_l p_{X,Y}(k,l) \qquad p_Y(l) = \sum_k p_{X,Y}(k,l)$$

	l_1	l_2	l_3	l_4	$p_X(k)$
k_1	0.1	0.1	0	0.2	0.4
k_2	0.05	0.05	0.1	0	0.2
k_3	0	0.1	0.2	0.1	0.4
$p_Y(l)$	0.15	0.25	0.3	0.3	

Computing Expectations

Expectation of a function: Let X and Y be random variables, and *g* be an arbitrary function of X and Y. Then

$$E[g(X,Y)] = \sum_{k}\sum_{l} g(k,l)\, p_{X,Y}(k,l)$$

Computing Expectations

Expectation of a function: Let X and Y be random variables, and g be an arbitrary function of X and Y. Then

$$E[g(X,Y)] = \sum_k \sum_l g(k,l)\, p_{X,Y}(k,l)$$

Linearity of expectation: Let X and Y be random variables. Then for any scalars a, b, and c,

$$\mathrm{E}[a\mathrm{X} + b\mathrm{Y} + c] = ???$$

Computing Expectations

Expectation of a function: Let X and Y be random variables, and *g* be an arbitrary function of X and Y. Then

$$E[g(X,Y)] = \sum_k \sum_l g(k,l)\, p_{X,Y}(k,l)$$

Linearity of expectation: Let X and Y be random variables. Then for any scalars *a*, *b*, and *c*,

$$\mathrm{E}[a\mathrm{X} + b\mathrm{Y} + c] = a\mathrm{E}[\mathrm{X}] + b\mathrm{E}[\mathrm{Y}] + c.$$

Three or More Random Variables

All of this extends to more than two random variables too.

Three or More Random Variables

All of this extends to more than two random variables too.

- Joint PMFs:

$$p_{X,Y,Z}(k, l, m) = P(X=k, Y=l, Z=m)$$
$$= P(\{X=k\} \cap \{Y=l\} \cap \{Z=m\})$$

Three or More Random Variables

All of this extends to more than two random variables too.

- Joint PMFs:

$$p_{X,Y,Z}(k, l, m) = P(X=k, Y=l, Z=m)$$
$$= P(\{X=k\} \cap \{Y=l\} \cap \{Z=m\})$$

- Expectation of a function:

$$E[g(X,Y,Z)] = \sum_k \sum_l \sum_m g(k,l,m)\, p_{X,Y,Z}(k,l,m)$$

etc.

Three or More Random Variables

An office decides to do a gift exchange for their annual holiday party. Everyone in the office purchases a gift and places it in a box. Each person is then given a random gift from the box. What is the expected number of people who get back the gift that they purchased?

Three or More Random Variables

An office decides to do a gift exchange for their annual holiday party. Everyone in the office purchases a gift and places it in a box. Each person is then given a random gift from the box. What is the expected number of people who get back the gift that they purchased?

Note that we used the fact that the expectation of a sum of random variables is equal to the sum of their expectations.

Independence is not required for this!

Conditioning

- The conditional PMF of X given Y is denoted $p_{X|Y}$. For any values k of X and l of Y,

$$p_{X|Y}(k \mid l) = P(X=k \mid Y=l) = P(\{X=k\} \mid \{Y=l\})$$

Conditioning

- The conditional PMF of X given Y is denoted $p_{X|Y}$. For any values k of X and l of Y,

$$p_{X|Y}(k \mid l) = P(X=k \mid Y=l) = P(\{X=k\}|\{Y=l\})$$

- We can compute this PMF using the definition of conditional probability

Conditioning

- We can compute conditional probabilities by normalizing the values in a particular row or column

	l_1	l_2	l_3	l_4
k_1	0.1	0.1	0	0.2
k_2	0.05	0.05	0.1	0
k_3	0	0.1	0.2	0.1

Conditioning

- We can compute conditional probabilities by normalizing the values in a particular row or column

	l_1	l_2	l_3	l_4
k_1	0.1	0.1	0	0.2
k_2	0.05	0.05	0.1	0
k_3	0	0.1	0.2	0.1

$$p_{X|Y}(k \mid l_2) = \frac{p_{X,Y}(k, l_2)}{p_Y(l_2)}$$

Conditioning

- We can compute conditional probabilities by normalizing the values in a particular row or column

	l_1	l_2	l_3	l_4
k_1	0.1	0.1	0	0.2
k_2	0.05	0.05	0.1	0
k_3	0	0.1	0.2	0.1

$$p_{X|Y}(k \mid l_2) = \frac{p_{X,Y}(k, l_2)}{p_Y(l_2)}$$

Conditioning

Conditional PMFs frequently give us a convenient method of calculating joint PMFs. Using the multiplication rule,

$$p_{X,Y}(k,l) = p_X(k)p_{Y|X}(l|k) = p_Y(l)p_{X|Y}(k|l)$$

Conditioning

Conditional PMFs frequently give us a convenient method of calculating joint PMFs. Using the multiplication rule,

$$p_{X,Y}(k,l) = p_X(k)p_{Y|X}(l|k) = p_Y(l)p_{X|Y}(k|l)$$

- Example: If my computer has a virus, which is true with probability 0.2, my anti-virus program detects it with probability 0.7. Otherwise, the program incorrectly identifies a virus with probability 0.1.

  Let X and Y be binary variables that are 1 if my computer has a virus and if the program says it has a virus respectively. What is $p_{X,Y}$?

Conditioning

We can condition random variables on events too. For a random variable X and event A associated with the same experiment,

$$p_{X|A}(k) = P(X=k \mid A) = P(\{X=k\} \mid A)$$

Conditioning

We can condition random variables on events too. For a random variable X and event A associated with the same experiment,

$$p_{X|A}(k) = P(X=k \mid A) = P(\{X=k\} \mid A)$$

Nothing fancy here. All of the same reasoning applies.

Conditioning

We can condition random variables on events too. For a random variable X and event A associated with the same experiment,

$$p_{X|A}(k) = P(X=k \mid A) = P(\{X=k\} \mid A)$$

Nothing fancy here. All of the same reasoning applies.

Remember: All of this is just notation. When in doubt, think about the underlying events.

Conditional Expectation

We can define a conditional version of expectation:

$$E[X \mid Y = l] = ???$$

Conditional Expectation

We can define a conditional version of expectation:

$$E[X \mid Y = l] = \sum_k k p_{X|Y}(k \mid l)$$

Conditional Expectation

We can define a conditional version of expectation:

$$E[X \mid Y = l] = \sum_k k p_{X \mid Y}(k \mid l)$$

For any function g:

$$E[g(X) \mid Y = l] = ???$$

Conditional Expectation

We can define a conditional version of expectation:

$$E[X \mid Y = l] = \sum_k k p_{X|Y}(k \mid l)$$

For any function *g*:

$$E[g(X) \mid Y = l] = \sum_k g(k) p_{X|Y}(k \mid l)$$

Conditional Expectation

We can define a conditional version of expectation:

$$E[X \mid Y = l] = \sum_k k p_{X|Y}(k \mid l)$$

For any function *g*:

$$E[g(X) \mid Y = l] = \sum_k g(k) p_{X|Y}(k \mid l)$$

Expectations can be conditioned on events A too.

Total Expectation Theorem

The Total Expectation Theorem says that for any random variables X and Y,

$$E[X] = \sum_{l} p_Y(l) E[X \mid Y = l]$$

Total Expectation Theorem

The Total Expectation Theorem says that for any random variables X and Y,

$$E[X] = \sum_{l} p_Y(l) E[X \mid Y = l]$$

More generally, for any disjoint events $A_1, \ldots, A_n$ that form a partition of Ω,

$$E[X] = \sum_{i=1}^{n} P(A_i) E[X \mid A_i]$$

Total Expectation Theorem

The Total Expectation Theorem says that for any random variables X and Y,

$$E[X] = \sum_{l} p_Y(l) E[X \mid Y = l]$$

More generally, for any disjoint events $A_1, \ldots, A_n$ that form a partition of Ω,

$$E[X] = \sum_{i=1}^{n} P(A_i) E[X \mid A_i]$$

We could prove this using the multiplication rule and the definition of conditional expectation – Try this at home!

Geometric Random Variables

Suppose we flip a coin with bias *p* repeatedly until it lands on heads. Let X be the number of times that we flip it.

Geometric Random Variables

Suppose we flip a coin with bias *p* repeatedly until it lands on heads. Let X be the number of times that we flip it.

$$p_X(k) = ???$$

Geometric Random Variables

Suppose we flip a coin with bias p repeatedly until it lands on heads. Let X be the number of times that we flip it.

$$p_X(k) = (1-p)^{(k-1)}\, p^k$$

Geometric Random Variables

Suppose we flip a coin with bias p repeatedly until it lands on heads. Let X be the number of times that we flip it.

$$p_X(k) = (1-p)^{(k-1)}\, p^k$$

We can use the total expectation theorem to compute the mean and variance…

Geometric Random Variables

Suppose we flip a coin with bias p repeatedly until it lands on heads. Let X be the number of times that we flip it.

$$p_X(k) = (1-p)^{(k-1)} p^k$$

We can use the total expectation theorem to compute the mean and variance…

(we only talked about the mean in class, but the calculation of variance uses similar ideas..)

Independence

- Recall that events A and B are independent if and only if

$$P(A \cap B) = P(A)\, P(B)$$

- If $P(B) > 0$, this condition is equivalent to

$$P(A \mid B) = P(A)$$

- If $P(A) > 0$, it is equivalent to

$$P(B \mid A) = P(B)$$

To prove or disprove independence, it is enough to prove or disprove any one of these conditions.

Independence

- X and Y are independent if and only if for all k and l

$$p_{X,Y}(k,l) = p_X(k)\, p_Y(l)$$

- This is equivalent to showing that for all k, l s.t. $p_Y(l)>0$

$$p_X(k) = p_{X|Y}(k \mid l)$$

- This is equivalent to showing that for all k, l s.t. $p_X(k)>0$

$$p_Y(l) = p_{Y|X}(l \mid k)$$

To prove or disprove independence, it is enough to prove or disprove any one of these conditions.

Independence

		Y			
		l_1	l_2	l_3	l_4
	k_1	0.05	0.15	0	0.2
X	k_2	0.025	0.075	0	0.1
	k_3	0.05	0.15	0	0.2

Are X and Y independent? How can you tell?

Independence

		Y			
		l_1	l_2	l_3	l_4
	k_1	0.05	0.15	0	0.2
X	k_2	0.025	0.075	0	0.1
	k_3	0.05	0.15	0	0.2

Columns are multiples of each other, which implies that $p_{X|Y}(k \mid l)$ is the same for all l

Independence

It is easy to show that if X and Y are independent, then

$$E[XY] = E[X]E[Y]$$

Independence

It is easy to show that if X and Y are independent, then

$$E[XY] = E[X]E[Y]$$

Note that this is not true in general – only holds when X and Y are independent!

Independence

The definition of independence can be extended in the natural way to sets of more than two random variables

$$p_{X,Y,Z}(k, l, m) = p_X(k)\, p_Y(l)\, p_Z(m)$$

Independence

The definition of independence can be extended in the natural way to sets of more than two random variables

$$p_{X,Y,Z}(k, l, m) = p_X(k)\, p_Y(l)\, p_Z(m)$$

If X, Y, and Z are independent, then any three random variables of the form f(X), g(Y), and h(Z) would be independent too.

One more useful fact…

If $X_1, X_2, \ldots, X_n$ are independent, then

$$\mathrm{var}(X_1 + X_2 + \ldots + X_n) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + \ldots + \mathrm{var}(X_n)$$

One more useful fact…

If $X_1, X_2, \ldots, X_n$ are independent, then

$$\mathrm{var}(X_1 + X_2 + \ldots + X_n) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + \ldots + \mathrm{var}(X_n)$$

This also is not true in general – need independence!

(we stopped here in class, but in the next lecture, we will see how to use this fact about the variance of sums of independent random variables to answer questions like the one on the next slide…)

The Sample Mean

Suppose we would like to estimate the president's approval rating. We ask n random voters whether or not they approve of the president, and use the fraction of voters who say that they approve as our estimate.

How accurate is our estimate as a function of n?