Lecture 8. Random variables and random vectors. Joint and marginal distributions. Expectation and variance.

YULIA R. GEL

CS/SE/STAT 3341 Probability and Statistics in Computer Science and Software Engineering

February 7, 2017

Motivation: A wheel of Fortune

Suppose you spin a wheel of fortune many times. At each spin, one of the numbers $m_1, m_2, \ldots, m_n$ comes up with corresponding probability $p_1, p_2, \ldots, p_n$, and this is your monetary reward from that spin.

What is the amount of money that you "expect" to get "per spin"?

The terms "expect" and "per spin" are a little ambiguous, but here is a reasonable interpretation.

Motivation: A wheel of Fortune

Suppose that you spin the wheel k times, and that k_i is the number of times that the outcome is m_i. Then, the total amount received is $m_1k_1 + m_2k_2 + \ldots + m_nk_n$. The amount received per spin is

$$M = \frac{m_1k_1 + m_2k_2 + \ldots + m_nk_n}{k}.$$

If the number of spins k is very large, and if we are willing to

interpret probabilities as relative frequencies, it is reasonable to anticipate that m_i comes up a fraction of times that is roughly equal to p_i:

$$p_i \approx \frac{k_i}{k}, \quad i = 1, \ldots, n.$$

Thus, the amount of money per spin that you expect to receive is

$$M = \frac{m_1k_1 + m_2k_2 + \ldots + m_nk_n}{k} \approx m_1p_1 + m_2p_2 + \ldots + m_np_n.$$

Expectation: Definition

Motivated by this example, we now introduce the notion of expectation.

Definition. We define the **expected value** (also called the **expectation** or the **mean**, or **averaged value**) of a random variable X, with PMF $P(x)$, by

$$E(X) = \sum_{x} xP(x).$$

The expected value is often denoted by μ.

Examples: Coins

Consider two independent coin tosses, each with a 0.75 probability of a head, and let X be the number of heads obtained. Its PMF is

Then the PMF of X is:

$$P(x) = \begin{cases} 0.25 \times 0.25, \text{ if } x = 0 \\ 0.25 \times 0.75 + 0.75 \times 0.25, \quad \text{if } x = 1 \\ 0.75 \times 0.75, \quad \text{if } x = 2 \end{cases}$$

Let us calculate its expected value:

$$\begin{aligned} \mu_X &= E(X) \\ &= 0 \times (0.25 \times 0.25) + 1 \times (0.25 \times 0.75 + 0.75 \times 0.25) \\ &+ 2 \times (0.75 \times 0.75) \\ &= \frac{3}{2}. \end{aligned}$$

Expectation: Properties

Let X and Y to be random variables, and a, b and c be non-random (deterministic). Then

1. $E(aX + bY + c) = aE(X) + bE(Y) + c$,
2. if X and Y are *independent* random variables, then $E(XY) = E(X)E(Y)$.

Variance, Covariance, Standard Deviation and Correlation: Definitions

Definitions:

- **Variance**, denoted by $\mathrm{var}(\mathrm{X})$, is defined as the expected value of the random variable $(X - E(X))^2$, i.e.

$$\mathrm{var}(\mathrm{X}) = \mathrm{E}(\mathrm{X} - \mathrm{E}(\mathrm{X}))^2.$$

  The variance provides a measure of dispersion (or spread) of X around its mean. For discrete r.v.

$$\mathrm{var}(\mathrm{X}) = \mathrm{E}(\mathrm{X}-\mathrm{E}(\mathrm{X}))^2 = \sum_{\mathrm{x}}(\mathrm{x}-\mathrm{E}(\mathrm{X}))^2\mathrm{P}(\mathrm{x}) = \sum_{\mathrm{x}}(\mathrm{x}-\mu)^2\mathrm{P}(\mathrm{x}).$$

Variance, Covariance, Standard Deviation and Correlation: Definitions

Definitions:

- Another measure of dispersion is the **standard deviation** of X, denoted by σ_X and defined as

$$\sigma_X = \sqrt{\mathrm{var(X)}} = \sqrt{E(X - E(X))^2}.$$

- Linear interrelation between X and Y can be summarized via **covariance**, i.e.

$$\sigma_{X,Y} = \mathrm{cov(X, Y)} = \mathrm{E(X - E(X))(Y - E(Y))}.$$

- However, covariance is not unit free. Instead, in practice we more often use **correlation**:

$$\rho = \mathrm{cor(X, Y)} = \frac{\mathrm{cov(X, Y)}}{\sqrt{\mathrm{var(X)var(Y)}}}.$$

Examples: Alice and STAT3341

If the weather is good, which happens with probability 0.6, Alice walks the 2 miles to class at a speed 5 miles/hr, otherwise she rides a bus at speed 30 miles/hr. What is the expected time to get to class?

Solution: Define the discrete r.v. X to take the value (2/5) hr with probability 0.6 and (2/30)hr with probability 0.4. The expected value of X

$$E(X) = 2/5 \times 0.6 + 2/30 \times 0.4 = 4/15 \ \ hour.$$

Now, variance of X:

$$\begin{aligned}
E(X-E(X))^2 &= \sum_x (x-E(X))^2 P(x) \\
&= \sum_x (x^2 - 2E(X)x + E(X)^2)P(x) \\
&= \sum_x (x^2 P(x) - 2E(X)xP(x) + E(X)^2 P(x)) \\
&= \sum_x (x^2 P(x) - 2E(X)\sum_x xP(x) + E(X)^2 \sum_x P(x) \\
&= EX^2 - 2(E(X))^2 + E(X)^2 = EX^2 - (E(X))^2 \\
&= [(2/5)^2 \times 0.6 + (2/30)^2 \times 0.4] - (4/15)^2 \\
&= 0.02667 hours^2.
\end{aligned}$$

The respective standard deviation **in hours** rather than **in hours**^{2} is $\sqrt{0.02667 hr^2} = 0.1633$.

Motivation

Often, we need to consider the relationship between two or more events:

- It is cloudy and windy.
- A cat purring and being groomed.
- Getting a flu and high fever.

Joint distributions allow us to reason about the relationship between multiple events.

Definitions

1. If X and Y are random variables, then the pair (X, Y) is a **random vector**.
2. Distribution of (X, Y) is called a **joint distribution** of X and Y. The joint PMF of X and Y is

$$P(x, y) = \mathbf{P}\{(X, Y) = (x, y)\} = \mathbf{P}\{(X = x \cap Y = y)\}.$$

3. Individual distribution of X and Y are called **marginal distributions** of X and Y. Marginalisation refers to the process of "removing" the inuence of one or more events from a probability.
4. Two random variables X and Y are called **independent** if $P(x, y) = P_X(x) \times P_Y(y)$.

Example: Corpus by Osborn and Keller, 2008

In information theory, corpus is a collection of text. Assume you have a corpus of a 100 words. You tabulate the words, their frequencies and probabilities in the corpus:

w	c(w)	P(w)	x	y
the	30	0.30	3	1
to	18	0.18	2	1
will	16	0.16	4	1
of	10	0.10	2	1
Earth	7	0.07	5	2
on	6	0.06	2	1
probe	4	0.04	5	2
some	3	0.03	4	2
Comet	3	0.03	5	2
CNN	3	0.03	3	0

Example: Corpus by Osborn and Keller, 2008

We can now define the following random variables:

- X is the length of the word;
- Y is number of vowels in the word.

Examples for probability mass functions:

- $P_X(5) = P(Earth) + P(probe) + P(Comet) = 0.14$;
- $P_Y(2) = P(Earth) + P(probe) + P(some) + P(Comet) = 0.17$.

Examples for cumulative distribution functions:

- $F_X(3) = P_X(2) + P_X(3) = 0.34 + 0.33 = 0.67$;
- $F_Y(1) = P_Y(0) + P_Y(1) = 0.03 + 0.80 = 0.83$.

Example: Corpus by Osborn and Keller, 2008

We can now model the relationship between word length (X and number of vowels Y.

Let $P(x, y) = P(X = x, Y = y)$.

Examples:

- $P(2, 1) = P(to) + P(of) + P(on) = 0.18 + 0.10 + 0.06 = 0.34$;
- $P(3, 0) = P(CNN) = 0.03$;
- $P(4, 3) = 0$.

Full joint distribution of (X, Y) is :

		2	3	4	5
	0	0	0.03	0	0
y	1	0.34	0.30	0.16	0
	2	0	0	0.03	0.14

Example: Corpus by Osborn and Keller, 2008

To calculate marginal distributions $P_X(x)$ and $P_Y(y)$, we can use the following formulas:

$$P_X(x) = \mathbf{P}\{X = x\} = \sum_y P_{X,Y}(x, y),$$

$$P_Y(y) = \mathbf{P}\{Y = y\} = \sum_x P_{X,Y}(x, y).$$

Thus, we get the marginal distribution of Y, $P_Y(y)$, from the full joint distribution of (X, Y) as follows:

		2	3	4	5	$P_Y(y) = \sum_x P_{X,Y}(x,y)$
	0	0	0.03	0	0	0.03
y	1	0.34	0.30	0.16	0	0.80
	2	0	0	0.03	0.14	0.17