Intro to Probability for Discrete Variables

BMI/CS 576

www.biostat.wisc.edu/bmi576/

Irene Ong

irene.ong@wisc.edu

Fall 2017

Definition of Probability

- *frequentist* interpretation: the probability of an event from a random experiment is proportion of the time events of the same kind will occur in the long run, when the experiment is repeated

- examples
  - the probability my flight to Chicago will be on time
  - the probability this ticket will win the lottery
  - the probability it will rain tomorrow

- always a number in the interval [0,1]
  0 means "never occurs"
  1 means "always occurs"

Sample Spaces

- *sample space*: a set of possible outcomes for some event

- examples
  - flight to Chicago: {on time, late}
  - lottery: {ticket 1 wins, ticket 2 wins,…,ticket *n* wins}
  - weather tomorrow:

    {rain, not rain} or

    {sun, rain, snow} or

    {sun, clouds, rain, snow, sleet} or…

Random Variables

- *random variable*: a function that maps the outcome of an experiment to a label (often a numerical value)
- example
  - X represents the outcome of my flight to Chicago
  - we write the probability of my flight being on time as $\Pr(X = \text{on-time})$
  - or when it's clear which variable we're referring to, we may use the shorthand $\Pr(\text{on-time})$

Notation

- UPPERCASE letters and Capitalized words denote random variables
- lowercase letters and uncapitalized words denote values
- we'll denote a particular value for a variable as follows

$$\Pr(X = x) \qquad \Pr(Fever = true)$$

- we'll also use the shorthand form

$$\Pr(x) \quad \text{for} \quad \Pr(X = x)$$

- for Boolean random variables, we'll use the shorthand

$$\Pr(fever) \;\; \text{for} \;\; \Pr(Fever = true)$$

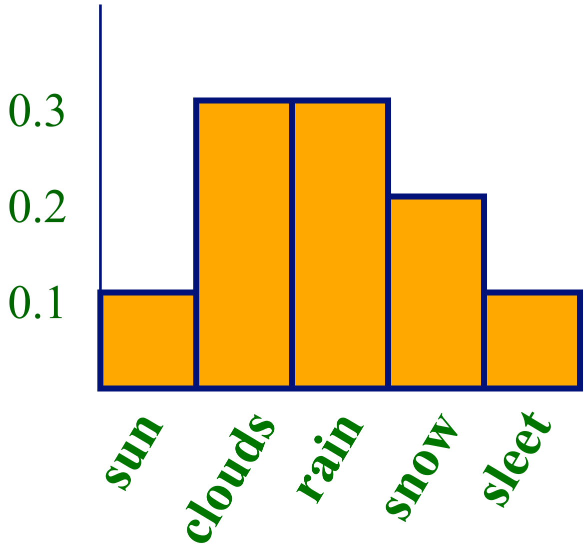
$$\Pr(\neg fever) \;\; \text{for} \;\; \Pr(Fever = false)$$

Probability Distributions

- if X is a random variable, the function given by $\Pr(X = x)$ for each x is the *probability distribution* of X
- requirements:

$$\Pr(x) \geq 0 \quad \text{for every } x$$

$$\sum_x \Pr(x) = 1$$



Joint Distributions

- *joint probability distribution*: the function given by $\Pr(X = x,\ Y = y)$
- read "X equals x and Y equals y"
- example

x, y	$\Pr(X = x,\ Y = y)$
sun, on-time	0.20
rain, on-time	0.20
snow, on-time	0.05
sun, late	0.10
rain, late	0.30
snow, late	0.15

probability that it's sunny and my flight is on time

Marginal Distributions

- the *marginal distribution* of X is defined by

$$\Pr(x) = \sum_{y} \Pr(x, y)$$

"the distribution of X ignoring other variables"

- this definition generalizes to more than two variables, e.g.

$$\Pr(x) = \sum_{y} \sum_{z} \Pr(x, y, z)$$

Marginal Distribution Example

joint distribution

x, y	$\Pr(X = x, Y = y)$
sun, on-time	0.20
rain, on-time	0.20
snow, on-time	0.05
sun, late	0.10
rain, late	0.30
snow, late	0.15

marginal distribution for X

x	$\Pr(X = x)$
sun	0.3
rain	0.5
snow	0.2

Conditional Distributions

- the *conditional distribution* of X given Y is defined as:

$$\Pr(X = x \mid Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)}$$

"the distribution of X given that we know Y"

Conditional Distribution Example

joint distribution

x, y	$\Pr(X = x,\ Y = y)$
sun, on-time	0.20
rain, on-time	0.20
snow, on-time	0.05
sun, late	0.10
rain, late	0.30
snow, late	0.15

conditional distribution for X given Y=on-time

x	$\Pr(X = x \mid Y = \textit{on-time})$
sun	0.20/0.45 = 0.444
rain	0.20/0.45 = 0.444
snow	0.05/0.45 = 0.111

Independence

- two random variables, *X* and *Y*, are *independent* if and only if

$$\Pr(x, y) = \Pr(x) \times \Pr(y) \quad \text{for all } x \text{ and } y$$

Independence Example #1

joint distribution

x, y	$\Pr(X = x, Y = y)$
sun, on-time	0.20
rain, on-time	0.20
snow, on-time	0.05
sun, late	0.10
rain, late	0.30
snow, late	0.15

marginal distributions

x	$\Pr(X = x)$
sun	0.3
rain	0.5
snow	0.2

y	$\Pr(Y = y)$
on-time	0.45
late	0.55

Are X and Y independent here? NO.

Independence Example #2

joint distribution

x, y	$\Pr(X = x, Y = y)$
sun, fly-United	0.27
rain, fly-United	0.45
snow, fly-United	0.18
sun, fly-Delta	0.03
rain, fly-Delta	0.05
snow, fly-Delta	0.02

marginal distributions

x	$\Pr(X = x)$
sun	0.3
rain	0.5
snow	0.2

y	$\Pr(Y = y)$
fly-United	0.9
fly-Delta	0.1

Are X and Y independent here? YES.

Conditional Independence

- two random variables *X* and *Y* are *conditionally independent* given *Z* if and only if

$$\Pr(X \mid Y, Z) = \Pr(X \mid Z)$$

"once you know the value of *Z*, knowing *Y* doesn't tell you anything about *X*"

- alternatively

$$\Pr(x, y \mid z) = \Pr(x \mid z) \times \Pr(y \mid z) \quad \text{for all } x, y, z$$

Conditional Independence Example

Flu	Fever	Vomit	Pr
true	true	true	0.04
true	true	false	0.04
true	false	true	0.01
true	false	false	0.01
false	true	true	0.009
false	true	false	0.081
false	false	true	0.081
false	false	false	0.729

Fever and Vomit are not independent: e.g. $\Pr(\mathit{fever}, \mathit{vomit}) \neq \Pr(\mathit{fever}) \times \Pr(\mathit{vomit})$

Fever and Vomit are conditionally independent given Flu:

$$\Pr(\mathit{fever}, \mathit{vomit} \mid \mathit{flu}) = \Pr(\mathit{fever} \mid \mathit{flu}) \times \Pr(\mathit{vomit} \mid \mathit{flu})$$

$$\Pr(\mathit{fever}, \mathit{vomit} \mid \neg \mathit{flu}) = \Pr(\mathit{fever} \mid \neg \mathit{flu}) \times \Pr(\mathit{vomit} \mid \neg \mathit{flu})$$

etc.

Chain Rule of Probability

- For two variables:

$$\Pr(X,Y) = \Pr(X \mid Y)P(Y)$$

- For three variables

$$\Pr(X,Y,Z) = \Pr(X \mid Y,Z)P(Y \mid Z)P(Z)$$

- etc.
- to see that this is true, note that

$$\Pr(X,Y,Z) = \frac{\Pr(X,Y,Z)}{P(Y,Z)}\frac{P(Y,Z)}{P(Z)}P(Z)$$

Bayes Theorem

$$\Pr(x \mid y) = \frac{\Pr(y \mid x)\Pr(x)}{\Pr(y)} = \frac{\Pr(y \mid x)\Pr(x)}{\sum_x \Pr(y \mid x)\Pr(x)}$$

- this theorem is extremely useful
- there are many cases when it is hard to estimate $\Pr(x \mid \mathrm{y})$ directly, but it's not too hard to estimate $\Pr(y \mid x)$ and $\Pr(x)$

Bayes Theorem Example

- MDs usually aren't good at estimating Pr(*Disorder* | *Symptom*)
- they're usually better at estimating Pr(*Symptom* | *Disorder*)
- if we can estimate Pr(*Fever* | *Flu*) and Pr(*Flu*) we can use Bayes' Theorem to get diagnosis

$$\Pr(flu \mid fever) = \frac{\Pr(fever \mid flu)\Pr(flu)}{\Pr(fever \mid flu)\Pr(flu) + \Pr(fever \mid \neg flu)\Pr(\neg flu)}$$

Expected Values

- the *expected value* of a random variable that takes on numerical values is defined as:

$$E[X] = \sum_x x \times \Pr(x)$$

this is the same thing as the *mean*

- we can also talk about the expected value of a function of a random variable (which is also a random variable)

$$E[g(X)] = \sum_x g(x) \times \Pr(x)$$

Expected Value Examples

$$E\left[Shoesize\right]=$$

$$5\times\Pr(Shoesize=5)+\ldots+14\times\Pr(Shoesize=14)$$

- Suppose each lottery ticket costs \$1 and the winning ticket pays out \$100. The probability that a particular ticket is the winning ticket is 0.001.

$$E\left[gain(Lottery)\right]=$$

$$gain(\text{winning})\Pr(\text{winning})+gain(\text{losing})\Pr(\text{losing})=$$

$$(\$100-\$1)\times 0.001-\$1\times 0.999=$$

$$-\$0.90$$

Linearity of Expectation

- An extremely useful aspect of expected values is the following identity

$$E[X+Y]=E[X]+E[Y]$$

- This holds even if X and Y are *not independent!*

Expected values of indicator random variables

- It is common to use *indicator* random variables

$$I_A = \begin{cases} 1 & \text{if event } A \text{ occurs,} \\ 0 & \text{otherwise} \end{cases}$$

- The expected value of such a variable is simply

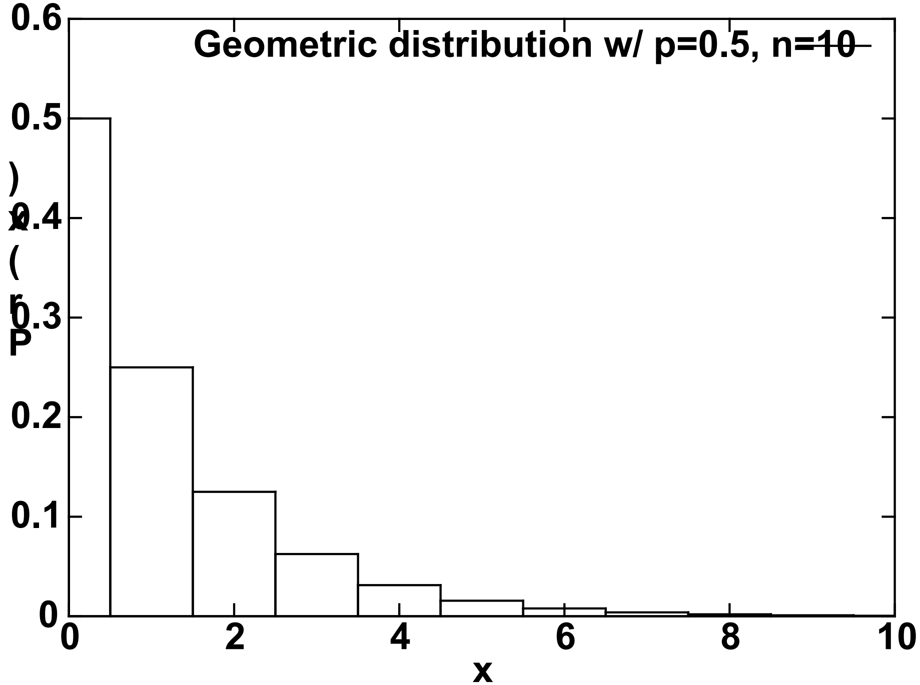
$$E[I_A] = 1 \times P(A) + 0 \times P(\neg A) = P(A)$$

The Geometric Distribution

- distribution over the number of trials before the first failure (with same probability of success p in each)

$$\Pr(x) = (1 - p)p^x$$

- e.g. the probability of x heads before the first tail

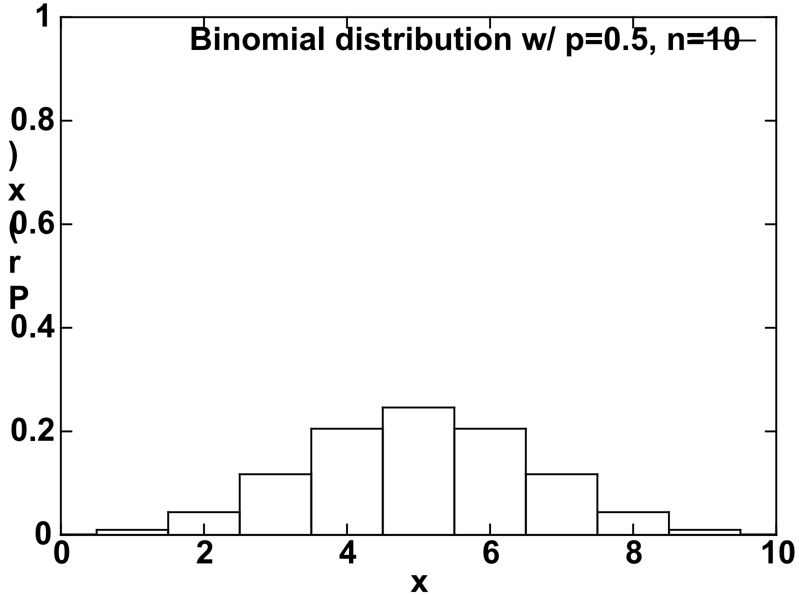


The Binomial Distribution

- distribution over the number of successes in a fixed number n of independent trials (with same probability of success p in each)

$$\Pr(x) = \binom{n}{x} p^{x} (1-p)^{n-x}$$

- e.g. the probability of x heads in n coin flips



The Multinomial Distribution

- k possible outcomes on each trial
- probability p_i for outcome x_i in each trial
- distribution over the number of occurrences x_i for each outcome in a fixed number n of independent trials

$$\Pr(x) = \frac{n!}{\prod_i (x_i!)} \prod_i p_i^{x_i}$$

For example, with $k = 6$ (e.g., a six-sided die) and $n = 30$:

$$\Pr([7,3,0,8,10,2]) = \frac{30!}{7! \times 3! \times 0! \times 8! \times 10! \times 2!} \left(p_1{}^7 p_2{}^3 p_3{}^0 p_4{}^8 p_5{}^{10} p_6{}^2 \right)$$

Continuous random variables

- When our outcome is a continuous number we need a continuous random variable
- Examples: Weight, Height
- We specify a density function for random variable *X* as

$$f(x) \geq 0$$

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

- Probabilities are specified over an interval to derive probability values

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

- Probability of taking on a single value is 0.

Continuous random variables contd

- To define a probability distribution for a continuous variable, we need to integrate *f(x)*

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

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

Examples of continuous distributions

- Gaussian distribution

- Exponential distribution

- Extreme Value distribution

Gaussian distribution

- Gaussian distribution

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$