CHAPTER 4: PROBABILITY AND PROBABILITY DISTRIBUTIONS

A, B: any two events ($\overline{A}$ is the complement of A)

Additive law

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Multiplicative law

$$P(A \text{ and } B) = P(A|B)P(B) = P(B|A)P(A)$$

Independence

A and B are independent if

$$P(A|B) = P(A)$$

Bayes' formula

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

This is Bayes' formula. Often the denominator probability must be obtained as a sum of joint probabilities, for instance, the event B can be written as the union of $(B \text{ and } A)$ and $(B \text{ and } \overline{A})$:

So

$$P(B) = P(B \text{ and } A) + P(B \text{ and } \overline{A})$$

$$= P(B|A)P(A) + P(B|\overline{A})P(\overline{A})$$

and

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})} \, .$$

ex. A drawer has 2 white socks and 2 blue socks. Professor reaches in and draw out 2 socks in succession, without replacement. $W_1 =$ "white on the first draw", $W_2 =$ "white on the second draw", $B_1 =$ "blue ..." etc. A tree diagram of the events, and the conditional probabilities, is:

W_2

$P(W_2|W_1) = \frac{1}{3}$

W_1

$P(B_2|W_1) = \frac{2}{3}$

B_2

$P(W_1) = \frac{1}{2}$

$P(B_1) = \frac{1}{2}$

W_2

$P(W_2|B_1) = \frac{2}{3}$

B_1

$P(B_2|B_1) = \frac{1}{3}$

B_2

Probability of two white socks:

$$P(W_1 \text{ and } W_2) = P(W_2|W_1)P(W_1) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$

Probability of two socks the same color:

$$P(W_1 \text{ and } W_2) + P(B_1 \text{ and } B_2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Probability that the *second* sock is blue:

$$P(B_2) = P(B_2|W_1)P(W_1) + P(B_2|B_1)P(B_1)$$

$$= \frac{2}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{2}$$

Probability that the *first* sock is white, given the *second* sock is blue:

$$P(W_1|B_2) = \frac{P(B_2|W_1)P(W_1)}{P(B_2)} \qquad \text{(Bayes' rule)}$$

$$= \frac{P(B_2|W_1)P(W_1)}{P(B_2|W_1)P(W_1) + P(B_2|B_1)P(B_1)}$$

$$\frac{\frac{2}{3} \cdot \frac{1}{2}}{\frac{2}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2}} = \frac{2}{3}$$

Discrete probability distributions

Random variable: a numerical outcome of a random experiment (usually denoted with an upper case letter, e.g. Y)

ex. # of blue socks
SAT of a randomly drawn student
democrats in a random sample of voters
1 day's growth (dry weight) of a plant

Probability distribution: collection of all possible outcomes of a random variable, and their associated probabilities (a particular outcome usually denoted with a lower case letter, e.g., y)

Discrete probability distribution: random variable has a finite or countably infinite number of states

1. Binomial distribution Y, a random variable, is the number of "successes" in n independent identical trials, in which each trial can be a success or a failure (possible outcomes are $y = 0, 1, 2, 3, ..., n$). For each trial, the probability of success is π, $0 < \pi < 1$ (not the pi from a circle). The binomial distribution has probabilities given by

$$P(Y = y) = \text{“}P(y)\text{”} = \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y}$$

for $y = 0, 1, 2, 3, ..., n$. These probabilities add to 1:

$$P(0) + P(1) + P(2) + \cdots + P(n) = 1.$$

The **expected value** or **mean** of the binomial random variable Y:

$$\mathrm{E}(Y) = \mu = 0 \cdot P(0) + 1 \cdot P(1) + \cdots + n \cdot P(n)$$

$$\underset{\text{ALA}}{=} \quad n\pi$$

Variance of the random variable Y:

$$\mathrm{V}(Y) = \sigma^2 = (0-\mu)^2 P(0) + (1-\mu)^2 P(1) + \cdots$$

$$+ (n-\mu)^2 P(n) \quad \underset{\text{ALA}}{=} \quad n\pi(1-\pi)$$

Common notation: $Y \sim$ binomial(n, π)

"Y has a binomial distribution with parameters n and π"

SAS:

The function RANBIN(seed, n, p) returns a binomial random variable with # trials n and success probability p (set seed = 0 and SAS will use the computer clock time as seed)

The function PROBBNML(p, n, x) computes the probability that an observation from a binomial(n, p) distribution will be less than or equal to x.

exercise (concept of mean and variance of a discrete distribution): suppose Y has a "rectangular distribution" given by

$$P(Y = y) = P(y) = \frac{1}{6}\,,$$

$$y = 1, 2, 3, 4, 5, 6$$

(distribution of the result of rolling a die). (a) Draw a picture of the probability distribution. (b) Calculate the expected value of Y. (c) Calculate the variance of Y.

2. Poisson distribution

$$P(Y = y) = P(y) = \frac{e^{-\mu}\mu^y}{y!}\,,$$

$$y = 0, 1, 2, 3, \ldots \quad .$$

Here $e = 2.71828...$ and μ is a parameter ($\mu > 0$). This is a distribution with positive probability on all the nonnegative integers:

$$P(0) + P(1) + P(2) + \cdots = \sum_{y=0}^{\infty} P(y) = 1.$$

Poisson distribution arises as a model of **rare events**:

radioactive decays in a unit of time
incoming cosmic rays in a unit of time
plant stems in a sample plot
steelhead caught in 1 hr
crimes reported in Moscow, ID in 1 day
car accidents reported in a stretch of US95 in 1 week

Note that **zero** is a possible outcome in the Poisson distribution. Some other properties:

$$\mathrm{E}(Y) = \mu$$

$$\mathrm{V}(Y) = \sigma^2 = \mu$$

Variance equals the mean in the Poisson distribution

$$P(Y = 0) = P(0) = e^{-\mu}$$

Notation: $Y \sim$ Poisson(μ)

True fact: suppose $Y \sim$ binomial(n, π), that is,

$$P(Y = y) = P(y) = \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y}$$

$$y = 0, 1, 2, \ldots, n.$$

Suppose n is **large** and π is **small**. Then

$$P(Y = y) \approx \frac{\mathrm{e}^{-\mu}\mu^y}{y!} \quad \text{where } \mu = n\pi$$

(Poisson approximation to the binomial)

SAS:

RANDPOI(seed, m) generates a Poisson random variable with mean m

POISSON(m, x) calculates $P(Y \leq \text{x})$, where $Y \sim \text{Poisson}(\mu)$

3. Multinomial distribution (a **multivariate** distribution)

k types

♦ ♦ ★ ♥ ♥	sample n
♦ ♥ ♦ ★	with replacement
★ ♦ ♥ ♥ ♥	

$\pi_1 =$ proportion of type 1 (★) in the urn
$\pi_2 =$ proportion of type 2 (♥) in the urn
$\vdots$
$\pi_k =$ propostion of type k (♦) in the urn

The π_j's are **constants** (parameters);

$$\pi_1 + \pi_2 + \cdots + \pi_k = 1.$$

$Y_1, Y_2, \ldots, Y_k$: **random variables**

$Y_1 =$ number of type 1 (★) in the **sample**
$Y_2 =$ number of type 2 (♥) in the **sample**
$\vdots$
$Y_k =$ number of type k (♦) in the **sample**

$$Y_1 + Y_2 + \cdots + Y_k = n.$$

The Y_j's are **dependent**: the value of one affects the others.

$$P(Y_1 = y_1 \text{ and } Y_2 = y_2 \text{ and } \cdots \text{ and } Y_k = y_k)$$

$$= \frac{n!}{y_1! y_2! \cdots y_k!} \pi_1^{y_1} \pi_2^{y_2} \cdots \pi_k^{y_k}$$

where y_1, y_2, ..., y_k are any nonnegative integers that add to n (all possible outcomes).

examples

$Y_1 =$ # democrats, $Y_2 =$ # republicans, $Y_3 =$ # greens, $Y_4 =$ # "other", in a random sample of n voters

$Y_1 =$ # genotype AA BB, $Y_2 =$ # AA Bb, $Y_3 =$ # AA bb, $Y_4 =$ # genotype Aa BB, ..., $Y_9 =$ # genotype aa bb in a random sample of n people

A population has n mice. Catch mice by live-trapping on two sampling occasions; uniquely identify each one caught.

$Y_1 =$ # mice caught in the first sample as well as the second
$Y_2 =$ # mice caught in the first sample but not in the second
$Y_3 =$ # mice not caught in the first sample but caught in the second
$Y_4 =$ # mice not captured in either sample (unobserved)
n: unknown