

EEB 600A, Lecture 2

Probability Theory and Discrete Distributions

Event, A

The complement of the Event A , A^c

State space: The possible events.

The Probability of an event $\Pr(A)$ satisfies the following

(i) $\Pr(A) \geq 0$, (ii) $\Pr(A) \leq 1$, (iii) Probabilities sum to one

Computing probabilities for multiple events

It may often be easier to compute the probability for the complement of a complex event, and then use

$$\Pr(A^c) = 1 - \Pr(A)$$

Example 1. Consider the toss of a coin, which is heads with probability p . What is the probability of seeing one or more heads in 100 flips? Well, one could compute this as

$$\sum_{i=1}^{100} \Pr(k \text{ heads}) = \Pr(\text{one head}) + \Pr(\text{two heads}) + \cdots + \Pr(\text{all heads})$$

or much more simply as

$$1 - \Pr(\text{no heads in 100 flips}) = 1 - (1 - p)^{100}$$

The And Rule : The Probability of two (or more) independent events

If the events A and B are independent, then

$$\Pr(A \text{ and } B) = \Pr(A) \cdot \Pr(B)$$

In other words, for independent events, the joint probability is given by the product of the individual probabilities,

$$\Pr(A_1 \text{ and } A_2 \text{ and } \dots \text{ and } A_n) = \prod_{i=1}^n \Pr(A_i)$$

Note that we usually use the shorthand of writing $\Pr(A \text{ and } B) = \Pr(A, B)$, the so called **joint probability** of events A and B .

Conditional Probabilities

If A and B are *not* independent, then knowing that (say) B has occurred provides information on the probability that A occurs. This **conditional probability** of event A given that event B is seen (A given B), is written $\Pr(A | B)$.

Likewise, the probability $\Pr(A)$ of Event A regardless of whether or not event B occurred, is called the **marginal probability** of A .

The joint, marginal, and conditional probabilities are related as follows:

$$\Pr(x, y) = \Pr(x | y) \cdot \Pr(y)$$

Hence, we can compute the conditional probability from

$$\Pr(x | y) = \frac{\Pr(x, y)}{\Pr(y)}$$

Example 2. Suppose the genotypes AA and Aa give a yellow flower, while aa is pink. Suppose the frequency of A is 0.75 in a population and that Hardy-Weinberg holds. What is the probability that a yellow flower is AA ? $\text{Freq}(AA) = 0.75^2 = 0.5625$ (9/16), while $\text{Freq}(Aa) = 2 \cdot (3/4) \cdot (1/4) = 0.375$ (3/8). Thus, $\Pr(\text{Yellow}) = \Pr(AA) + \Pr(Aa) = 0.9375$ (15/16), and

$$\Pr(AA | \text{yellow}) = \frac{\Pr(\text{yellow and } AA)}{\Pr(\text{yellow})} = \frac{\Pr(AA)}{\Pr(\text{yellow})} = \frac{9/16}{15/16} = 0.6$$

Baye's theorem. Suppose there are n possible outcomes (b_1, b_2, \dots, b_n) of a random variable that we cannot directly observe. Given the observed outcome of a correlated variable A , what is the probability of b_j ? From the definition of a conditional probability, $\Pr(b_j | A) = \Pr(b_j, A) / \Pr(A)$. We can decompose this further, by noting that $\Pr(b_j, A) = \Pr(b_j) \Pr(A | b_j)$ and $\Pr(A) = \sum_{i=1}^n \Pr(b_i) \Pr(A | b_i)$. Putting these together gives Bayes' theorem,

$$\Pr(b_j | A) = \frac{\Pr(b_j) \Pr(A | b_j)}{\Pr(A)} = \frac{\Pr(b_j) \Pr(A | b_j)}{\sum_{i=1}^n \Pr(b_i) \Pr(A | b_i)}$$

Example 3. Consider a genetic condition that causes all offspring to be female, and suppose the frequency of this disorder in the population is 0.01. Suppose you observe a family with seven girls (and no boys). What is the chance that they have the disorder? Here b_1 = disorder family, b_2 = normal family.

$$\Pr(7 \text{ girls} \mid \text{disorder}) = 1, \quad \Pr(7 \text{ girls} \mid \text{normal}) = (1/2)^7$$

$$\begin{aligned} \Pr(7 \text{ girls}) &= \Pr(7 \text{ girls} \mid \text{dis}) \cdot \Pr(\text{dis}) + \Pr(7 \text{ girls} \mid \text{nor}) \cdot \Pr(\text{nor}) \\ &= 1 \cdot 0.01 + (1/2)^7 \cdot 0.99 = 0.0178 \end{aligned}$$

Putting these together,

$$\Pr(\text{dis} \mid 7 \text{ girls}) = \frac{\Pr(\text{dis}) \cdot \Pr(7 \text{ girls} \mid \text{dis})}{\Pr(7 \text{ girls})} = \frac{0.01 \cdot 1}{0.0178} = 0.564$$

The Or Rule: the probability of a set of disjoint events. If the events A and B are disjoint (non-overlapping), then

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$$

If A and B overlap, then

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A, B)$$

Discrete Probability Distributions

For discrete events, the probability distribution of a random variable X is completely specified by $\Pr(X = x)$, where X is a random variable and x a particular outcome (for example, X is a roll of the dice, and $X = 5$ is the outcome of a roll of five).

Common Discrete Distributions

1. Bernoulli. This distribution is specified by p , the probability of a success. The outcomes are just $X = 0$ (failure) or $X = 1$ (success), with

$$P(x) = \begin{cases} 0 & \text{with probability } 1 - p \\ 1 & \text{with probability } p \end{cases}$$

For this distribution, the mean $\mu = p$ and the variance $\sigma^2 = p(1 - p)$. The Bernoulli serves as the building-block for other common distributions (binomial, geometric, negative binomial).

2. Binomial. The sum of n independent Bernoulli random variables. The parameters are the success probability p and the sample size n . The probability of x successes in n trials is given by

$$P(x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}, \quad \text{for } 0, 1, 2, \dots, n$$

Mean $\mu = np$, variance $\sigma^2 = np(1-p)$

3. Geometric. Specified by the probability of a success p . The number of trials required until the first success. The probability that the first success occurs on the x th trial is given by

$$P(x) = p(1-p)^{x-1}, \quad \text{for } 1, 2, \dots, \infty$$

Mean $\mu = (1-p)/p$, variance $\sigma^2 = (1-p)/p^2$. The probability of at least one success in k trials is thus given by

$$Pr(\text{at least one success}) = 1 - Pr(\text{no successes}) = 1 - (1-p)^k$$

4. Negative binomial. This is the generalization of the geometric for the number of trials until r successes occur. The distribution parameters are r and the success probability p . The probability of a total of x failures before r successes is

$$P(x) = \frac{(r+x-1)!}{(r-1)!x!} p^r (1-p)^x, \quad \text{for } 0, 1, 2, \dots, \infty$$

Mean $\mu = r(1-p)/p$, variance $\sigma^2 = r(1-p)/p^2$

5. Poisson. Specified by the parameter λ , the expected number of events. The poisson is a model constant risk, for example that chance of a success in a very small time interval Δt is $\lambda\Delta t$. Likewise, in ecology we often consider a constant chance of success not over time but rather over some area being measured, where for a very small area ΔA , the probability of a success is $\lambda\Delta A$

The probability tha x events occur is given by

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad \text{for } 0, 1, 2, \dots, \infty$$

In particular, the probability of no successes is given by

$$Pr(0) = e^{-\lambda}$$

The the Poisson, the mean and variance are the same, $\mu = \sigma^2 = \lambda$.

The Poisson is often used to approximate a Binomial distribution when the sample size n is large, but p is small. Since the expected number of successes $\lambda = np$, one can approximate the Binomial with n, p by a Poisson with parameter $\lambda = np$.

Example 4. The Poisson correction for sequence comparisons. Suppose we were examining the rate of evolution of a protein sequence. Two sequences are aligned from two species of interest, and we note that the sequences match at 40 of 100 sites. Assuming each mutation is different, what is the estimated per site rate of substitution? Note that two sites that differ could each have fixed different mutations, so simply using $60/100 = 0.60$ per site underestimates the true rate. If λ is the true rate, the probability that a site unremains unchanged is $e^{-\lambda}$, which we can estimate from the frequency of sites with no mutations. Hence

$$e^{-\lambda} = 0.4, \quad \text{or } \lambda = -\ln(0.4) = 0.916$$
