Using Densities

Conditional probability

If A and B are events in S, and $P(B) > 0$, then the *conditional probability* of A given B, written $P(A|B)$, is

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

Conditional probability example

$$S = \{1,2,3,4,5,6\}$$
$$A = \{2\}, P(A) = \frac{1}{6}$$
$$B = \{2\} \cup \{4\} \cup \{6\}, P(B) = \frac{1}{2}$$
$$A \cap B = \{2\}, P(A \cap B) = \frac{1}{6}$$
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}$$

Another example

Type	probability
Productive, talks	.4
Productive, quiet	.2
Useless, talks	.1
Useless, quiet	.3

Now suppose that the value of a productive RA is 1 and the value of an unproductive RA is zero and that the cost of an RA is 0.7 and that you want to maximize

$V = P(\text{Prod}) \times (1 - .7) + P(\text{Unprod}) \times (0 - .7)$

Insurance market

Consider a health insurance plan. If a person is sick, the plan will have to spend C. A healthy person costs nothing. The probability that a young person will be sick is p_y; for an old person p_O, where $p_o > p_y$. A fraction f of the population is young. People voluntarily buy insurance if the premium is equal to or below their expected cost of illness. The insurance company prices the plan so as to break even.

A. Suppose participation is mandatory. What will be the premium?
B. Suppose participation is voluntary. What will be the premium?

Bayes rule

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

Equating these gives

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

Bayes rule

Let $A_1, A_2, \ldots$ be a partition of the sample space, and let B be any set. Then $\forall i$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{\infty} P(B|A_i)P(A_i)}$$

Somewhat true facts about tests for tuberculosis

- Tuberculosis is a serious disease that is on the increase.
- Incidence is about 2 people per 10,000 (varies across countries and time)
- Inexpensive tests are pretty accurate.
  - Probability of a false positive is 2 out of 100
  - Probability of a false negative is 1 out of 100

Example of tuberculosis test

- B is outcome of tuberculosis test
  - $B \in \{yes, no\}$
- A is whether patient has tuberculosis
  - $A \in \{yes, no\}$
  - $p(A = yes) = 2/10{,}000$
- $p(B = yes|A = yes) = .99$
- $p(B = yes|A = no) = 0.02$

What is $p(A = yes|B = yes)$, i.e. tuberculosis if test is positive?

Example of tuberculosis test, cont.

$$P(\text{tuber}|\text{pos}) = \frac{P(\text{pos}|\text{tuber}) \times P(\text{tuber})}{P(\text{pos}|\text{tuber})P(\text{tuber}) + P(\text{pos}|\text{~tuber})P(\text{~tuber})}$$

$$P(\text{tuber}|\text{pos}) = \frac{.99 \times .0002}{.99 \times .0002 + .02 \times .9998} = 0.0098$$

Independence

Two events, A and B are *statistically independent* if

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

Intuition:

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

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

Example of not independent

Type	probability
Productive, talks	.4
Productive, quiet	.2
Useless, talks	.1
Useless, quiet	.3

$$P(\text{Prod} \cap \text{Talks}) = .4$$
$$P(\text{Prod}) = .6$$
$$P(\text{Talks}) = .5$$
$$.4 \neq .6 \times .5$$

Further implications of independence

If A and B are independent events, then the following pairs are also independent:

a) A and B^c

b) A^c and B

c) A^c and B^c

Random variable

A *random variable* is a function from a sample space S into the real numbers.

The *cumulative distribution function* or *cdf* of a random variable X, denotes by $F_X(x)$, is defined by

$$F_X(x) = P_X(X \leq x), \forall x$$

Sum of Two Dice CDF

$$F(1) = 0$$
$$F(2) = \frac{1}{36}$$
$$F(3) = \frac{1}{36} + \frac{2}{36}$$
$$\vdots$$
$$F(12) = 1$$

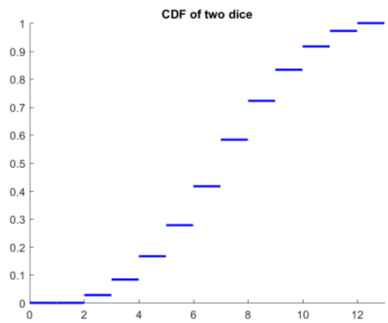
CDF properties

The function $F(x)$ is a cdf iff the following three conditions hold:

a) $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to \infty} F(x) = 1.$

b) $F(x)$ *is a nondecreasing function of* x.

c) $F(x)$ *is right-continuous: i.e.,* $\forall x_0, \lim_{x \downarrow x_0} F(x) = F(x_0).$



Logistic cdf

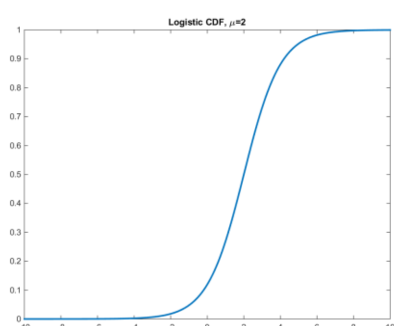
$$F_X(x) = \frac{1}{1 + e^{-\frac{(x-\mu)}{s}}}, s > 0$$

a) $\lim_{x \to -\infty} F(x) = 0$ since $\lim_{x \to -\infty} e^{-x} = \infty$ and $\lim_{x \to \infty} F(x) = 1$ since $\lim_{x \to \infty} e^{-x} = 0$.

b) $F(x)$ *is a nondecreasing function of* x since

$$\frac{d}{dx} F_X(x) = \frac{e^{-\frac{(x-\mu)}{s}}}{s\left(1+e^{-\frac{(x-\mu)}{s}}\right)^2} > 0$$

c) Is continuous and therefore right-continuous.



Identically distributed random variables

The random variables X *and* Y *are identically distributed if* $F_X(x) = F_Y(x) \forall x$

Density functions

The *probability mass function* for a discrete distribution is

$$f_X(x) = P(X = x)$$
$$F_X(x) = \sum_{k=-\infty}^{x} f_X(x)$$

The *probability density function* a continuous distribution is

$$f_X(x) = \frac{d}{dx} F_X(x)$$
$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx$$

Logistic density, $\mu = 0, s = 1$

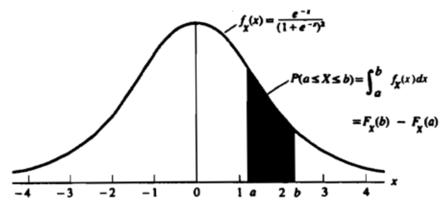


Figure 1.6.1. *Area under logistic curve*

Expected value

$$\mathrm{E}(x) = \sum_x x \cdot f(x)$$
$$\mathrm{E}(x) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

$$\mathrm{E}(bet) = -1 \times p + 1 \times (1-p)$$
$$= 1 \times (1 - 2p)$$

Variance

$$\mathrm{var}(x) = \sum_x (x - \mathrm{E}(x))^2 \cdot f(x)$$
$$\mathrm{var}(x) = \int_{-\infty}^{\infty} (x - \mathrm{E}(x))^2 \cdot f(x)dx$$
$$standard\ deviation \equiv \sqrt{var}$$

$$\mathrm{E}(bet) = B(1-2p)$$

$\mathrm{var}(bet)$
$= \left(-B - B(1-2p)\right)^2 p + \left(B - B(1-2p)\right)^2 (1-p)$

If $p = \frac{1}{2}, \mathrm{var}(bet) = B^2, std\ dev(bet) = B$

Bernoulli

$$f(x) = \begin{cases} p, & x = 1 \\ 1-p, & x = 0 \\ 0, & otherwise \end{cases}$$
$$f(x|p) = p^x(1-p)^{1-x}, x \in \{0,1\}$$

$$\mathrm{E}(x) = 1 \times p + 0 \times (1-p) = p$$
$$\mathrm{var}(x) = (1-p)^2 \times p + (0-p)^2 \times (1-p)$$
$$= p(1-p)$$

Support

The support of a distribution is the set of values for which the probability is positive.

The support of the Bernoulli is $\{0,1\}$.

Binomial distribution

$$f(x|n,p) = \binom{n}{x} p^x(1-p)^{n-x}$$

$$\mathrm{E}(x) = np$$
$$\mathrm{var}(x) = np(1-p)$$

Random outcome of a test statistic

Consider a test statistic that rejects a true null hypothesis a fraction $p = 0.05$ of the time. Suppose we have $n = 20$ independent applications of this test. Draw a plot of the pdf of the number of rejections (an application of the binomial distribution). What is the probability of more than 1 rejection?

