Probability Review

Survey Sampling
Statistics 4234/5234
Fall 2018

September 6, 2018

Example 1: Flip a coin 3 times, the sample space is

$$\Omega = \{HHH,\ HHT,\ HTH,\ HTT,\ THH,\ THT,\ TTH,\ TTT\}$$

Definition: The *sample space* for a random experiment is the set of possible outcomes.

Assume Ω is finite, $\Omega = \{\omega_1, \ldots, \omega_k\}$.

Associated with each outcome ω_i is a probability p_i satisfying

$$p_i \geq 0 \text{ for } i = 1, \ldots, k \quad \text{and} \quad \sum_{i=1}^{k} p_i = 1$$

Definition: A collection of outcomes (any subset of the sample space) is called an *event*. The probability of an event is the sum of the probabilities of the outcomes that make up that event.

Example 1: $p_i = \frac{1}{8}$ for $i = 1, \ldots, 8$. Define the event A to be "exactly two heads" so $A = \{HHT,\ HTH,\ THH\}$ and $P(A) = \frac{3}{8}$.

Simple random sampling with replacement (SRSwR)

N balls in urn, sample one ball n times, replacing the ball in the urn between each draw.

There are N^n posible ordered samples (permutations), each equally likely.

Example 2: Population of $N = 5$ units, sample $n = 2$ with replacement. The probability that unit 5 is in the sample is

$$P\left(\{15, 25, 35, 45, 51, 52, 53, 54, 55\}\right) = \frac{9}{25} = 0.36$$

Simple random sampling without replacement (SRS)

N balls in urn, draw n of them at random.

There are

$$N \times (N-1) \times \cdots \times (N-n+1) = \frac{N!}{(N-n)!}$$

possible ordered samples (permutations).

Ignoring the order we have

$$\frac{N!}{n!(N-n)!} = \binom{N}{n}$$

possible samples, equally likely.

Example 2: With $N = 5$ and $n = 2$ there are $\binom{5}{2} = 10$ possible samples. The probability that unit 5 is in our sample is

$$P\left(\{15, 25, 35, 45\}\right) = \frac{4}{10} = 0.40$$

Example 3: An urn has 5 black balls and 3 red ones, we will draw 4 at random. Then

$$P(\text{no red}) = \frac{5 \times 4 \times 3 \times 2}{8 \times 7 \times 6 \times 5} = \frac{\binom{5}{4}\binom{3}{0}}{\binom{8}{4}} = \frac{1}{14}$$

and

$$P(\text{one red}) = \frac{\binom{5}{3}\binom{3}{1}}{\binom{8}{4}} = \frac{3}{7} \quad \text{and} \quad P(\text{two reds}) = \frac{\binom{5}{2}\binom{3}{2}}{\binom{8}{4}} = \frac{3}{7}$$

Random variables

A *random variable* assigns a numeric value to each outcome,

$$X : \Omega \to \mathbb{R}$$

The set of possible values and their probabilities is the *probability distribution* of the random variable.

Example 3: Urn contains 5 black and 3 red balls, pick 4 at random, let $X =$ the number of reds.

x	0	1	2	3
$P(X = x)$	1/14	6/14	6/14	1/14

Definition: The *expected value* of the random variable X is

$$E(X) = \sum_x x P(X = x)$$

Example 3: $E(X) = 1.5$

Proposition: For any random variable X and function $g(\cdot)$, the mean of the random variable $g(X)$ is

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

Definition: The *variance* of the random variable X is

$$V(X) = E\left[(X - EX)^2\right]$$

Example 3: $V(X) = 0.5357$

Proposition: An alternative expression for the variance is

$$V(X) = E(X^2) - (EX)^2$$

Joint distributions

Example 4: Let the random variables (X, Y) have the joint probability distribution given by the following table of $P(X = x, Y = y)$.

	y		
x	1	2	3
1	1/6	1/6	1/6
2	1/12	0	1/12
3	0	1/3	0

Proposition: Given the pair of random variables (X, Y), and a function $g(\cdot, \cdot)$, the expected value of the random variable $g(X, Y)$ is

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) P(X = x, Y = y)$$

Example 4: Find $E(X)$ and $E(Y)$ and $E(XY)$.

OK.

$$E(X) = 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{3}\right) = \frac{11}{6}$$

and $E(Y) = 2$, by inspection, and

$$E(XY) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{6}\right) + 6\left(\frac{5}{12}\right) = \frac{11}{3}$$

Definition: The *covariance* between the random variables X and Y is

$$\text{Cov}(X,Y) = E\left[(X - EX)(Y - EY)\right]$$

and the *correlation* is

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{V(X)\cdot V(Y)}}$$

Proposition: $\text{Cov}(X, Y) = E(XY) - (EX)(EY)$.

Definition: The random variables X and Y are *independent* if and only if

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

for all x and y.

Proposition: If X and Y are independent then $\text{Cov}(X, Y) = 0$.

The converse of this propostion is not true (see Example 4).

Conditional probability

Definition: The *conditional probability* of B given A, where $P(A) > 0$, is

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

Proposition: If the events $A_1, \ldots, A_k$ form a partition of the sample space, that is if $A_1 \cup \cdots \cup A_k = \Omega$ and $A_i \cap A_j = \varnothing$ for all $i \neq j$, then

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

Example 5: Suppose we have three urns, $U_1 = \{4, 6, 7, 9\}$ and $U_2 = \{6, 8\}$ and $U_3 = \{5\}$. We first randomly select an urn, then randomly select a number from that urn.

Then

$$P(4) = P(U_1)P(4|U_1) + 0 + 0 = \left(\frac{1}{3}\right)\left(\frac{1}{4}\right) = \frac{1}{12}$$

and

$$P(5) = 0 + 0 + P(U_3)P(5|U_3) = \frac{1}{3}(1) = \frac{1}{3}$$

and

$$\begin{aligned} P(6) &= P(U_1)P(6|U_1) + P(U_2)P(6|U_2) \\ &= \left(\frac{1}{3}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right) = \frac{1}{4} \end{aligned}$$

Conditional expectation

Given discrete random variables X and Y, and a number x with $P(X = x) > 0$, the *conditional distribution* of Y given $X = x$ is defined by

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

The *conditional expected value* of Y given $X = x$ is

$$E(Y|X = x) = \sum_y yP(Y = y|X = x)$$

and the *conditional variance* of Y given $X = x$ is

$$V(Y|X = x) = E\left\{[Y - E(Y|X = x)]^2 \,|X = x\right\}$$

For each value of x with $P(X = x) > 0$ we can define the functions

$$g(x) = E(Y|X = x) \quad \text{and} \quad h(x) = V(Y|X = x)$$

and thus have defined the random variables

$$g(X) = E(Y|X) \quad \text{and} \quad h(X) = V(Y|X)$$

which have some interesting properties:

1. $E(Y) = E[E(Y|X)]$

2. $V(Y) = V[E(Y|X)] + E[V(Y|X)]$

Example 5: Define the random variables X and Y so that X indicates the urn selected (1, 2 or 3), and Y equals the number drawn from that urn. It is straightforward to show that Y is distributed as

y	4	5	6	7	8	9
$P(Y=y)$	1/12	4/12	3/12	1/12	2/12	1/12

and compute

$$E(Y) = 6.17 \qquad \text{and} \qquad V(Y) = 2.14$$

The conditional distributions $P(Y = y|X = x)$ for $x = 1, 2, 3$ are given in the following table, from which are easily computed the conditional means and variances $E(Y|X = x)$ and $V(Y|X = x)$:

x	4	5	6	7	8	9	E	V
1	1/4	0	1/4	1/4	0	1/4	6.50	3.25
2	0	0	1/2	0	1/2	0	7.00	1.00
3	0	1	0	0	0	0	1.00	0.00

Then we have

$$\begin{aligned} E[E(Y|X)] &= \sum_{x=1}^{3} E(Y|X = x)P(X = x) \\ &= \frac{1}{3}(6.5 + 7.0 + 5.0) = 6.17 \end{aligned}$$

confirming Property 1 above.

Also

$$E[V(Y|X)] = \sum_{x=1}^{3} V(Y|X=x)P(X=x)$$
$$= \frac{1}{3}(3.25 + 1.00 + 0.00) = 1.4167$$

and

$$V[E(Y|X)] = \sum_{x=1}^{3} [E(Y|X=x)]^2 \, P(X=x) - (EY)^2$$
$$= \frac{1}{3}(6.5^2 + 7.0^2 + 5.0^2) - 6.17^2 = 0.7222$$

and thus

$$V[E(Y|X)] + E[V(Y|X)] = 2.14$$

confirming Property 2.