Statistical Methods MATH2715 info

Assessment

- In course: 15 % R progr. + 5 % other assignments ; 80 % Exam.

10 Credits

- 22 lectures; 10 Workshops; 5 handouts ($\sim$ every other week).

Resources

- Mathematical Statistics and Data Analysis - 3rd ed. (by J. A. Rice);
- Introduction to Statistics - Online Edition -D.M.Lane et al.
- http://www1.maths.leeds.ac.uk/statistics/R/Rintro.pdf;
- https://www.datacamp.com/courses/free-introduction-to-r.

Random Variables

Why?

- In an experiment we are often interested in some value associated with an event as opposed to the actual event itself;
- f.e. tossing a coin three times: we may not be interested in the actual head-tail sequence that results but more in the number of heads that occur.
- These quantities of interest are *random numbers* called *random variables*.

Random Variables

Why?

- In an experiment we are often interested in some value associated with an event as opposed to the actual event itself;
- f.e. tossing a coin three times: we may not be interested in the actual head-tail sequence that results but more in the number of heads that occur.
- These quantities of interest are *random numbers* called *random variables*.

Definition

Definition. A *random variable* X on a sample space Ω is a real-valued function that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.

Random Variables

Random Variable

Definition. A *random variable* X on a sample space Ω is a real-valued function that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.

Notations

We use capital letters such as X for random variables. The notation $X \leq x$ is shorthand for the event $\{\omega \in \Omega | X(\omega) \leq x\}$.

Random Variables

Discrete Random Variable

Definition. A *discrete random variable* a r.v. that can take only a finite or at most a countably infinite number of values.

Continuous Random Variable

Definition. A *continuous random variable* a r.v. that can take on a *continuum* of values.

Discrete Random Variables

Probability of an event

For a discrete random variable (d.r.v.) X and a real value a, the event "X=a" is the set of outcomes in Ω for which the random variable assumes the value a, i.e., $X = a \equiv \{\omega \in \Omega | X(\omega) = a\}$. The probability of this event is denoted by

$$\Pr[X = a] = \sum_{\omega \in \Omega : X(\omega) = a} \Pr[\omega]$$

Probability mass function

The *frequency* or *probability mass* function (PMF) of a d.r.v. X gives the probabilitites for the different possible values of X. Thus, if x is a value that X can assume, the probability mass of X, $p_X(x)$, and is s.t.

$$p_X(x) = \Pr[X = x] \textit{ and } \sum_x p_X(x) = \sum_x \Pr[X = x] = 1$$

Cumulative Distribution

The **cumulative distribution function** (cdf) is a non decreasing function F defined as:

$$F(x) = P(X \leq x), \forall - \infty < x < \infty.$$

and satisfies:

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

Independence

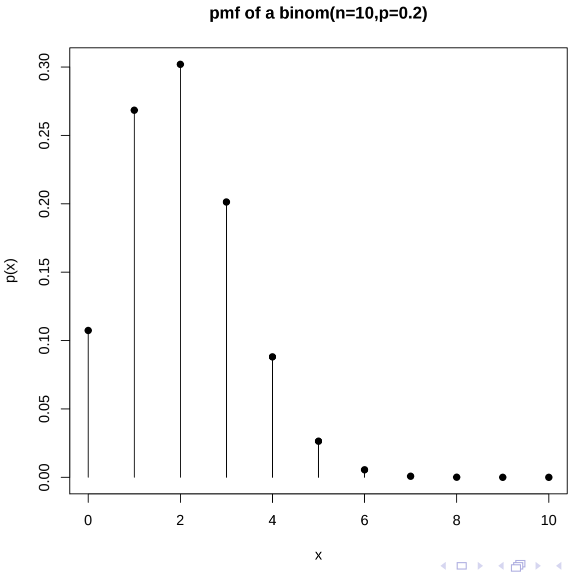
Two random variables, X and Y , are said to be independent if every event expressible in terms of X alone is independent of every other event expressible in terms of Y alone. In particular,

$$P(X \leq x \text{ and } Y \leq y) = P(X \leq x)P(Y \leq y).$$

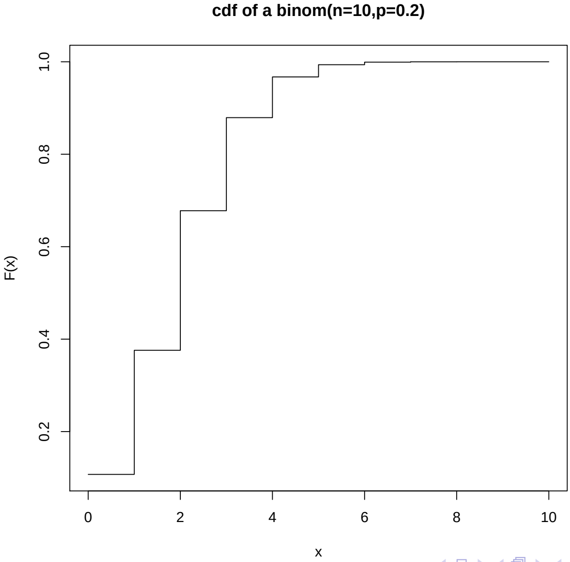
If the random variables are discrete,$X[Y]$ taking on possible values $x_i[y_i]$ then:

$$P(X = x_i \text{ and } Y = y_j) = P(X = x_i)P(Y = y_j).$$

A discrete mass function and the corresponding cdf



A discrete mass function and the corresponding cdf



R implementation

```
pdf('Figures/binom_density.pdf')
x <- 0:10
y <- dbinom(x,10,.2)
plot(x,y,type="h",xlab = "x",ylab="p(x)",main="pmf of
points(x,y,pch=19)
dev.off()

pdf('Figures/binom_cdf.pdf')
x <- 0:10
y <- pbinom(x,10,.2)
plot(x,y,type="s",xlab = "x",ylab="F(x)",main="cdf of
dev.off()
```

Discrete Random Variables: An example.

Consider the experiment of tossing three fair coins. The sequesnce of h and t is observed. What is the sample space?

Discrete Random Variables: An example.

Consider the experiment of tossing three fair coins. The sequesnce of h and t is observed. What is the sample space?

Sample space

$$\Omega = \{hhh, hht, htt, hth, ttt, tth, thh, tht\}$$

Can we define a random variable on Ω?

Discrete Random Variables: An example.

Consider the experiment of tossing three fair coins. The sequesnce of h and t is observed. What is the sample space?

Sample space

$$\Omega = \{hhh, hht, htt, hth, ttt, tth, thh, tht\}$$

Can we define a random variable on Ω?
Let X be the random variable that denotes the number of heads that results.

PMF of X

The possible value of X are $0, 1, 2$ *and* 3. Each of the outcomes in Ω has probability $1/8$, hence the PMF of X is:

$$p_X(x) = \begin{cases} 1/8 & \text{if } x = 0 \text{ or } x = 3 \\ 3/8 & \text{otherwise} \end{cases}$$

Bernoulli Random Variables

A Bernoulli r.v. takes only two values: 0 and 1. It's frequency function is:

$$p(x) = \begin{cases} p & \text{if } x = 1, \\ 1 - p & \text{if } x = 0 \\ 0 & \text{otherwise.} \end{cases}$$

A useful representation is:

$$p(x) = \begin{cases} p^x(1-p)^{1-x} & \text{if } x = 1 \text{ or } x = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Bernoulli Random Variables: interpretation

Let $A \subset \Omega$ be an event in a sample space ?. Let

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases}$$

X is an **indicator random variable**, often denoted by $X(\omega) = 1_A(\omega)$, which takes the value one if event A happens, zero otherwise.

Bernoulli r.v. as indicators

Bernoulli random variables often represent *success* vs. *failure* of an experiment and hence usually occourr as indicators.

Binomial Distribution

Assume that:

- an experiment is performed n times;
- each experiment performed (*trial*) independently of the others;
- assume that each experiment results in a *success* with probability p (i.e., each experiment is described by a Bernoulli r.v. Y_j).

Definition

The random variable $X = \sum_j^n Y_j$, denoting the number of successes in the n independent Bernoulli trials, has a **binomial distribution**:

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Assume:

- a sequence of independent Bernoulli trials is performed;
- there is no upper bound on the number of trials;

Definition

The random variable X, denoting the number of trials that must be performed until a *success* occurs, has a geometric distribution:

$$p(k) = P(X = k) = p(1-p)^{k-1} = pq^{k-1}\ ,\ k = 1, 2, \ldots .$$

NT

$$\sum_{k=1}^{\infty} p(1-p)^{k-1} = p\sum_{j=0}^{\infty} p(1-p)^{j} = 1.$$

we are using the **geometric series** sum $S = \sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$ where $|r| < 1$.

Negative Binomial Distribution

Same assumptions as the Geometric distribution:

- a sequence of independent Bernoulli trials is performed;
- there is no upper bound on the number of trials.

Definition

The r.v. X, denoting the number of trials required **until the r-th** success (where r is some given integer),has a Negative Binomial Distribution. The event $\{X = k\}$ happens when in the first $k-1$ trials there were exactly $r-1$ successes and on the $k-th$ trial there was also a success. Hence,

$$p(k) = P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

Poisson Distribution

The Poisson Distribution can be derived as the limit of the binomial distribution:

- consider a Binomial distribution with very large n and very small p;
- Let $\lambda = pn$;
- Let $n \to \infty$ and $p \to 0$ s.t. λ remains constant.

The limiting distribution is called the Poisson Distribution. Indeed:

Poisson frequency function

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{1}$$

- The PMF of a random variable, X, provides us with the probabilities of all possible values of X.
- It would be desirable to summarize this distribution into a representative number that is also easy to compute.
- This *might* be accomplished by the *expectation* of a random variable!

Expectation of a discrete r.v.

The *expectation* of a discrete random variable X, denoted by $E[X]$, is given by

$$E[X] = \sum_k k p_x(k) = \sum_k k \Pr[X = k]$$

Expectation Example

In the experiment of tossing three fair coins, we considered the r.v. X denoting the number of heads that result and we computed the PMF given below:

$$p_X(x) = \begin{cases} 1/8 & \text{if } x = 0 \text{ or } x = 3 \\ 3/8 & \text{otherwise} \end{cases}$$

Expectation of X

In our running example, in expectation the number of heads is given by

$$E[X] = 0 \times \frac{1}{8} + 3 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} = \frac{3}{2}$$

As seen from the example, the expectation of a random variable may not be a valid value of the random variable.

Expectation Example: Roll a die

Example

When we roll a die what is the result in expectation?

Solution.

Let X be the random variable that denotes the result of a single roll of dice. The PMF for X is given by

$$p_x(k) = \frac{1}{6}, k = 1, 2, 3, 4, 5, 6.$$

The expectation of X is given by

$$E[X] = \sum_{x=1}^{6} p_x(x) \cdot x = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

Roll two dice, what is the expected value of the sum?

Let S be the random variable denoting the sum. The PMF and $E[S]$ are:

$$p_S(x) = \begin{cases} \frac{1}{36}, x = 2, 12 \\ \frac{2}{36}, x = 3, 11 \\ \frac{3}{36}, x = 4, 10 \\ \frac{4}{36}, x = 5, 9 \\ \frac{5}{36}, x = 6, 8 \\ \frac{6}{36}, x = 7 \end{cases}$$

$$\begin{aligned} E[S] &= \sum_{x=2}^{12} p_S(x) \cdot x \\ &= \frac{1}{36} \times 2 + \frac{2}{36} \times 3 + \frac{3}{36} \times 4 + \frac{4}{36} \times 4 + \frac{5}{36} \times 6 + \frac{6}{36} \times 7 + \\ &\quad \frac{5}{36} \times 8 + \frac{4}{36} \times 9 + \frac{3}{36} \times 10 + \frac{2}{36} \times 11 + \frac{1}{36} \times 12 \\ &= \frac{252}{36} = 7 \end{aligned}$$

Linearity of Expectation

The linearity property of the expectation implies that the expectation of the sum of random variables equals the sum of their expectations.

Theorem

For any finite collection of random variables $X_1, X_2, \ldots, X_n$,

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Proof

We will prove the statement for two random variables X and Y. The general claim can be proven using induction.

$$\begin{aligned}
E[X+Y] &= \sum_i \sum_j (i+j)\Pr[X=i \cap Y=j] \\
&= \sum_i \sum_j (i\Pr[X=i \cap Y=j] + j\Pr[X=i \cap Y=j]) \\
&= \sum_i \sum_j i\Pr[X=i \cap Y=j] + \sum_i \sum_j j\Pr[X=i \cap Y=j] \\
&= \sum_i i \sum_j \Pr[X=i \cap Y=j] + \sum_j j \sum_i \Pr[X=i \cap Y=j] \\
&= \sum_i i\Pr[X=i] + \sum_j j\Pr[Y=j] \\
&= E[X]+E[Y]
\end{aligned}$$

Lemma

For any constant c and discrete random variable X:

$$E[cX] = cE[X]$$

Proof.

The lemma clearly holds for $c = 0$. For $c \neq 0$

$$\begin{aligned} E[cX] &= \sum_j j \Pr[cX = j] \\ &= c \sum_j (j/c) \Pr[X = j/c] \\ &= c \sum_k k \Pr[X = k] \\ &= cE[X] \end{aligned}$$

Exercise

- **A**. Using linearity of expectation calculate the expected value of the sum of the numbers obtained when two dice are rolled.

Solution

A. Let X_1 and X_2 denote the random variables that denote the result when die 1 and die 2 are rolled respectively. We want to calculate $E[X_1 + X_2]$. By linearity of expectation

$$
\begin{aligned}
E[X_1 + X_2] &= E[X_1] + E[X_2] \\
&= \frac{1}{6}(1+2+3+4+5+6) + \frac{1}{6}(1+2+3+4+5+6) \\
&= 3.5 + 3.5 = 7
\end{aligned}
$$