Bivariate Distributions

Concepts similar to Univariate r.v.

- Joint distribution
- Expectation

New concepts related with Bivariate r.v.

- Marginal distribution
- Conditional distribution
- Independent r.v.
- Covariance, correlation coefficient

Independence of Random Variables

- Review the concepts of independence for two events

- The random variables x and y are **independent** if and only if for all values of x and y

$$p(x,y) = p_x(x)p_y(y) \quad x \text{ and } y \text{ are discrete}$$

$$f(x,y) = f_x(x)f_y(y) \quad x \text{ and } y \text{ are continuous}$$

- That is, the following are equivalent to the random variables x and y being independent:

$$\begin{aligned} f_{y|x}(y \mid x) &= f_y(y) \\ f_{x|y}(x \mid y) &= f_x(x) \end{aligned}$$

  Same for discrete probability functions.

- If x and y are independent, then

$$\mathbb{E}(xy) = \mathbb{E}(x)\mathbb{E}(y)$$

Covariance

- How two variables, say x and y, vary together? For example, do they cluster along some line?

- Can quantity $\mathbb{E}[xy]$ be a reasonable measure here?

- The **covariance** is defined as

$$\text{Cov}(x, y) = \mathbb{E}[(x - \mu_x)(y - \mu_y)]$$

- Note that if $x = y$, this reduces to the variance σ_x^2

- An equivalent expression for covariance

$$\begin{aligned} \text{Cov}(x, y) &= \mathbb{E}[xy - \mu_x y - \mu_y x + \mu_x \mu_y] \\ &= \mathbb{E}(xy) - \mu_x \mu_y - \mu_y \mu_x + \mu_x \mu_y \\ &= \mathbb{E}(xy) - \mu_x \mu_y \end{aligned}$$

- $\text{Cov}(ax + b, cy + d) = ac\text{Cov}(x, y)$ where a, b, c, d are constants.

- If x and y are independent, then

$$\text{Cov}(x, y) = 0.$$

Correlation

- It turns out that when the covariance is normalized by dividing by the product of σ_x and σ_y, then its value will always be in the interval $[-1, 1]$,

$$-1 \leq \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y} \leq 1$$

- This "normalized covariance" is called the **coefficient of correlation**, and usually denoted by ρ:

$$\rho \equiv \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y}$$

- $\rho = 1$ and $\rho = -1$ imply deterministic linear relationships between x and y, the former with a positive slope and the latter with a negative slope. $\rho = 0$ implies no **linear** relationship between x and y.

True or False?

- The correlation is near -1 when the points are tightly packed along a line with negative slope.

- Correlations near 0 indicate the scatter plot shows at least one curve.

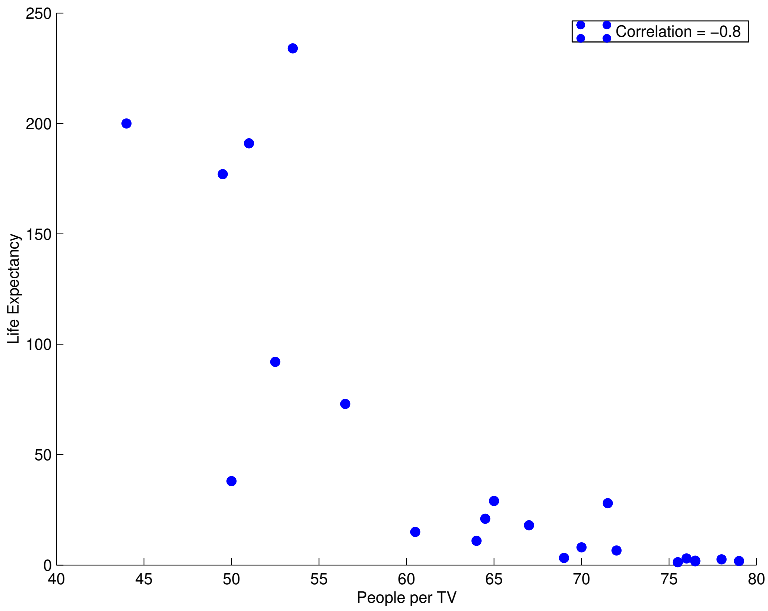
- Let x be the temperature in NYC and y be temperature in LA in degrees Fahrenheit. Changing y to degrees Celsius changes the value of the correlation.

- Correlation is an appropriate measure for non-linear relationships too.

- A newspaper article contains a quote from a psychologist, who says, “The evidence indicates the correlation between the research productivity and teaching rating of faculty members is close to zero.” The paper reports this as “The professor said that good researchers tend to be poor teachers, and vice versa.”

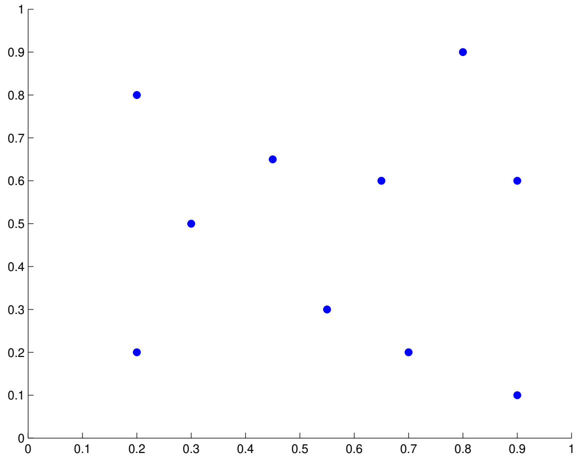
Important Facts about Correlation

- Correlation is not causation.

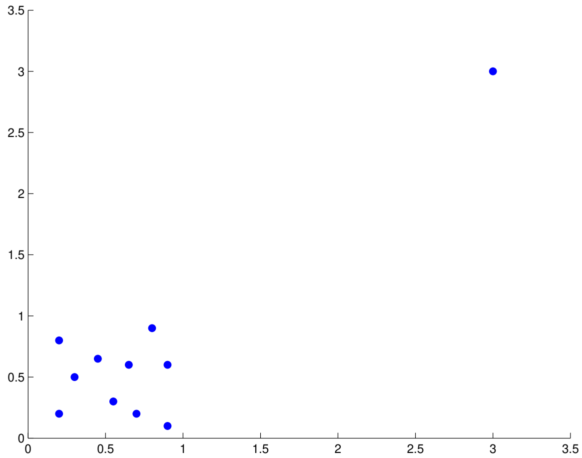


Scatter plot of life expectancy of population and number of people per TV for 22 countries (1991 data)

- Correlations can be strongly affected by outliers.



Correlation $= -0.08$



Correlation $= 0.89$

Correlation Matrix

- How to examine many correlations simultaneously

- The **correlation matrix** displays correlations for all pairs of variables.

```
>> cov(lifetv)
ans =
   1.0e+03 *

    0.1180   -0.6596
   -0.6596    5.7055

>> corrcoef(lifetv)
ans =

    1.0000   -0.8038
   -0.8038    1.0000
```

Independent vs Uncorrelated

- Independence implies $\rho = 0$: if x and y are independent, then

$$\begin{aligned} \text{Cov}(x, y) &= \mathbb{E}[(x - \mu_x)(y - \mu_y)] \\ &= \mathbb{E}(x - \mu_x)\mathbb{E}(y - \mu_y) \\ &= 0 \end{aligned}$$

- Uncorrelated random variables are NOT necessarily independent. But if they are **uncorrelated** and **normal**, then they are independent.

Contents in Chap 7

- Distributions of Functions of r.v.
    - CDF method (Sec 7.3)

- Simulation (Sec 7.4)

- Distribution of Sampling Statistics
    - Definitions (Sec 7.2)
    - Central Limiting Theorem (Sec 7.5, 7.6)

The CDF method

- Let w be a function of random variables.
- Find the probability $P(w \leq w_0)$, which is (dropping the subscript 0) is equal to $F(w)$.
- $f_w(w) = dF(w)/dw$.

Example: $y \sim N(\mu, \sigma^2)$. Find the distribution for $w = ay + b$ where a and b are constants.

Answer: $w \sim N(a\mu + b, a^2\sigma^2)$, that is, a linear transformation of a normal r.v. is still a normal r.v.

Example 7.4 $x, y \sim U(0,1)$ (independent). Find the density function for the sum $w = x + y$.

- What's the range of w?
- What's the joint density function for (x, y)?
- Use CDF method. First find $F(w_0) = P(w \leq w_0)$.
  - When $w_0 \leq 1$

  - When $w_0 > 1$

Distribution for CDF $F(y)$

Let y be a continuous r.v. with density function $f(y)$ and CDF $F(y)$.

Q: What's the distribution for $w = F(y)$?

A: $w \sim$ Uniform(0,1).

Using CDF method:

- There is a one-to-one correspondence between y values and w values.

$$w \le w_0 \Longrightarrow y \le y_0$$

- The CDF for w is

$$P(w \le w_0) = P(y \le y_0) = F(y_0) = w_0.$$

  Therefore the CDF for w is equal to

$$F(w) = w.$$

- $f(w) = dF(w)/dw = 1$ for $0 \le w \le 1$.

Example: Height of Plants

Without an automated irrigation system, the height of plants two weeks after germination is normally distributed with a mean of 2.5 centimeters and a standard deviation of 0.5 centimeters.

It is reasonable to assume that with an automated irrigation system, the height of plants two weeks after germination is also normally distributed.

How to guess the mean?

- A **statistic** is any quantity whose value can be calculated from sample data. For example, sample mean.

- A statistic is random variable.

- A statistic's distribution is often called **sampling distribution** to emphasize that it describes how the staistic varies in value across all samples.

- The **standard error** of a statistic is the standard deviation of its sampling distribution.

Some Facts of Normal

- Linear transformation of a normal is still a normal: If $y \sim N(\mu, \sigma^2)$, then

$$ay + b \sim N(a\mu + b, a^2\sigma^2).$$

- Linear combination of normals is still a normal: If $y_i \sim N(\mu_i, \sigma_i^2)$, then

$$a_1 y_1 + a_2 y_2 \cdots a_m y_m \sim N(\mu, \sigma^2)$$

where

$$\begin{aligned} \mu &= a_1\mu_1 + a_2\mu_2 + \ldots a_m\mu_m \\ \sigma^2 &= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \ldots a_m^2\sigma_m^2 \\ &+ \sum_{i \neq j} 2a_i a_j \text{cov}(y_i, y_j) \end{aligned}$$

- If $y_1, y_2, \ldots, y_m$ are independently drawn from $N(\mu, \sigma^2)$, what's the distribution of the sample mean $\bar{y}$?

Simulation

Example 7.6&7.7 Simulate the distribution of sample mean

$$\bar{y} = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5}$$

where y_i's are independently distributed as (1) Norm(0,1); (2) Unif(0,1); (3) Exp(1).

How to Approximate a distribution by simulation

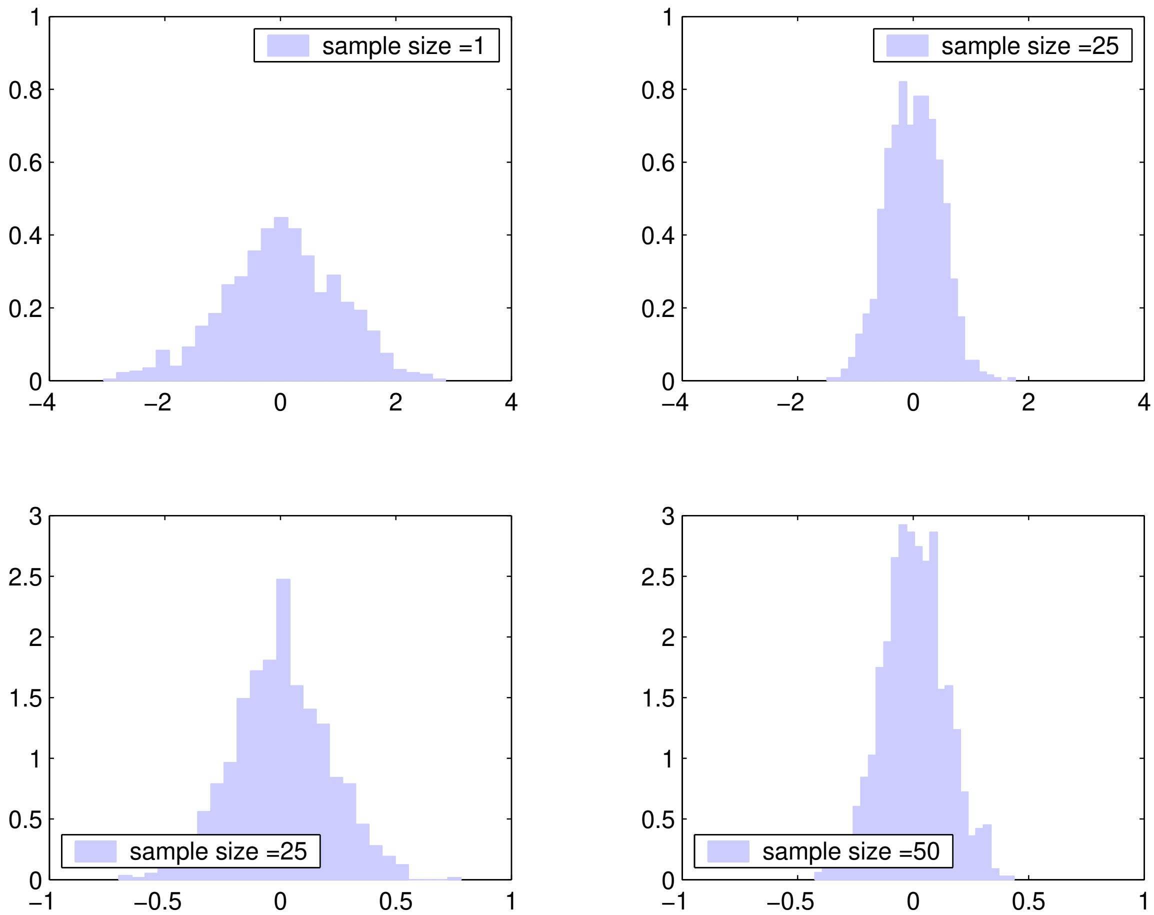
(1) generate data from the distribution

(2) draw histogram with height equal to (relative freq / bin width)

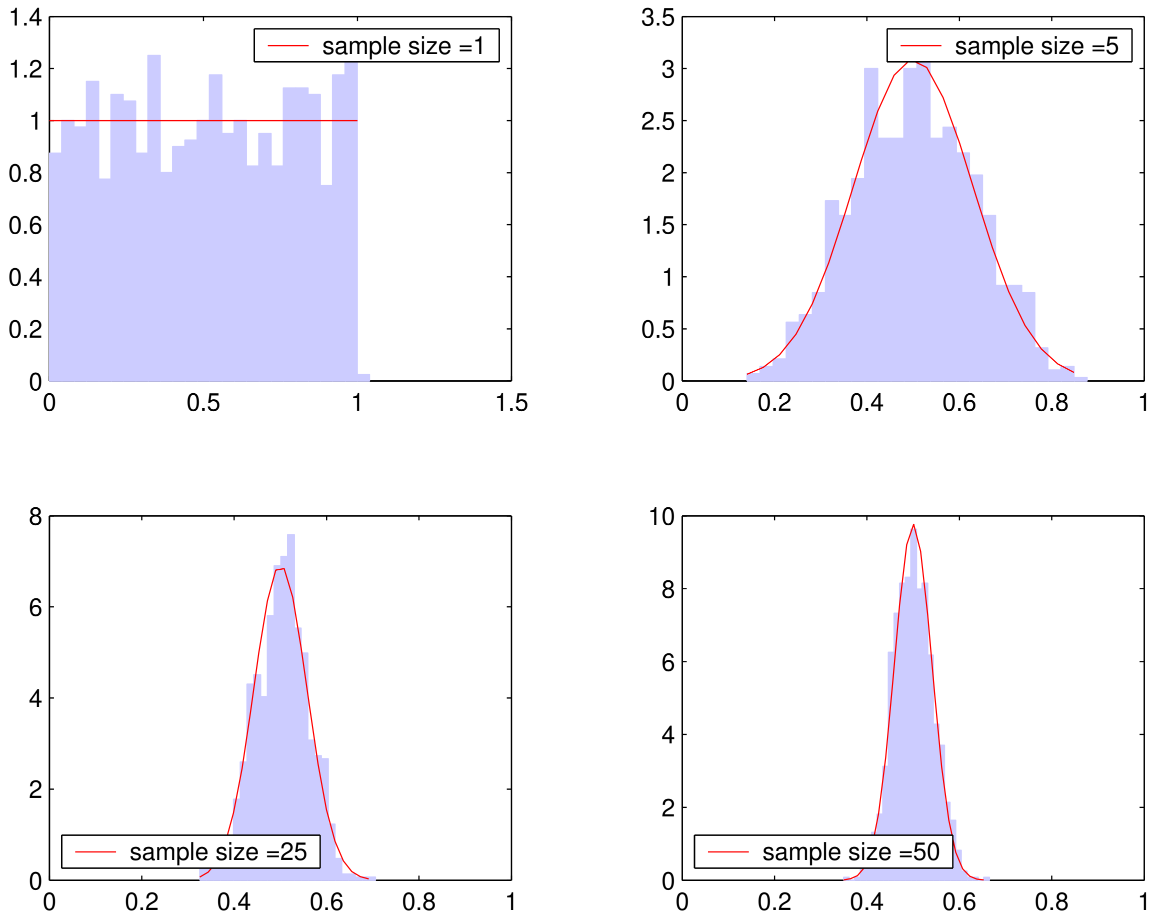
(3) the histogram can be used to approximate the density

```
function f=sdist(data, bin)
  n = length(data);
  binsize = range(data)/bin;
  edg= min(data):binsize:max(data);
  [count, junk] = histc(data, edg);
  h = bar(edg, count./(n*binsize), 'histc');

m=5; n=1000;
y =unifrnd(0,1,n,m)
    0.5126    0.6116    0.5014    0.9178    0.1305
    0.2317    0.4498    0.7678    0.0448    0.6383
    0.3946    0.4529    0.4626    0.9537    0.6609
    0.3848    0.3681    0.8608    0.3324    0.6767
    .....

mean(y,2)
    0.5348
    0.4265
    0.5849
    0.5246
    .....

sdist(mean(y,2), 25);
```

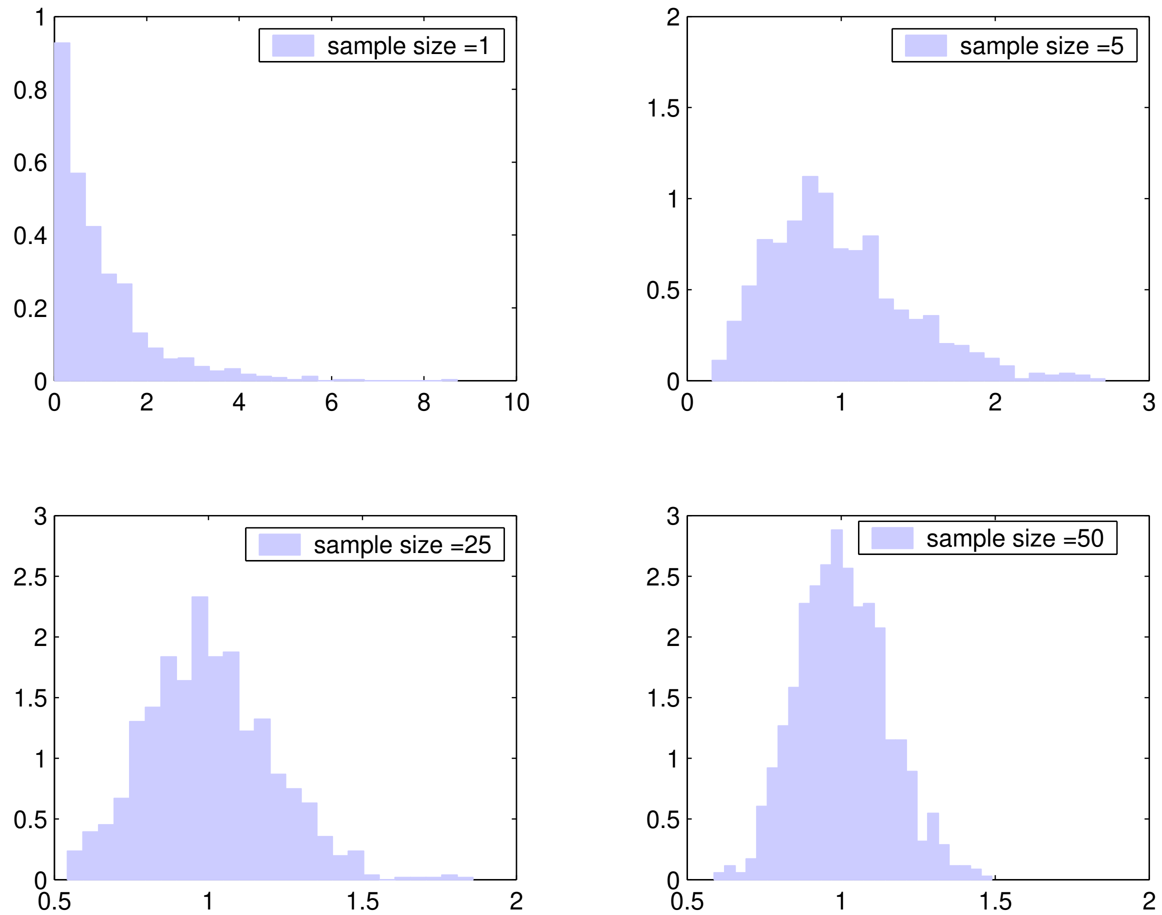
Distribution of a random sample of 100 sample mean from a Norm(0,1) distribution.



Distribution of a random sample of 100 sample mean from a Unif(0,1) distribution.



Distribution of a random sample of 100 sample mean from a Exp(1) distribution.



Central Limit Theorem

- $y_1, \dots y_n$ are drawn from a distribution **independently** with finite mean μ and variance σ^2, then

$$\mathbb{E}\bar{y} = \mu, \quad \mathrm{Var}(\bar{y}) = \sigma^2/n.$$

- **Central Limit Theorem** When n is sufficiently large, $\bar{y}$ can be approximated by a normal distribution with mean μ and variance σ^2/n, i.e.,

$$\frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- The sampling distribution of a sum of random variables

$$\frac{\sum_{i=1}^{n} y_i - n\mu}{n\sigma^2} \sim N(0,1)$$