Class 5 - Statistics and Probability (part I)

Pedram Jahangiry

Jon M. Huntsman School of Business
Utah State University

Probability review

✓ Experiment: Any procedure that can be infinitely repeated and has a well-defined set of outcomes.

✓ Random variable: A variable that takes on numerical values and the outcome is determined by experiment.

- Discrete random variable: Number of heads in a coin-flipping experiment, number of times observing a 4 when throwing a dice, letter GPA, number of bedrooms in a house, number of stocks in a portfolio and …
- Continuous random variable: Height, Weight, time, distance, numerical GPA, GDP, int rate, prices and …

✓ Probability Density Function (PDF): summarizes the probabilities of outcomes!

✓ Cumulative Distribution Function (CDF): summarizes the cumulative probabilities of outcomes!

Probability review

✓ Conditional probabilities: A conditional probability is the probability of an event, given some other event has already occurred

$$P(Y|X) = \frac{P(Y \cap X)}{P(X)}$$

✓ **Independence**

$$P(Y \cap X) = P(Y)P(X) \rightarrow P(Y|X) = P(Y)$$

If Y and X are independent random variables, then knowledge of the value taken on by X tells us nothing about the probability that Y takes on various values and vice versa.

Features of probability distributions

✓ Measures of central tendency:

1. Expected value
   - If X is a discrete random variable: $E(X) = \sum_{i=1}^{k} x_i f(x_i)$
   - If X is a continuous random variable: $E(X) = \int_{-\infty}^{\infty} x f(x)$
2. Median
   - Median m, is a value such that one half of the area under the pdf is to the left of m and one-half of the area is to the right of m.

❑ Median is less sensitive than the average to extreme values (outliers). Ex: median income or median housing prices (why?)

❑ If X has a symmetric distribution about the mean, then the expected value and median are equal.

❑ Properties of Expected values:

1. For any constant c, $E(c) = c$
2. For any constant a and b, $\mathrm{E}(aX + b) = aE(X) + b$
3. The <u>Expected value of the sum</u> is equal to the <u>sum of expected values</u>.

Features of probability distributions

✓ Conditional Expectation:

$$E(Y|X) = \sum y_i f(Y|X)$$

❑ Properties of Conditional expectation:

1. **Mean Independence**: If X and Y are independent, then $\boldsymbol{E(Y|X) = E(Y)}$

   This means that the expected value of Y given X does not depend on X.

2. $E[f(X)|X] = f(X)$

   Intuitively, this simply means that if we know X, then we also know f(X)

3. For functions $f(X)$ and $g(X)$

$$E[\{ f(X)Y + g(X)\} \mid X] = f(X)\, E(Y|X) + g(X)$$

Features of probability distributions

✓ Measures of dispersion:

1. Variance: How far, on average, is X from its mean!
2. Standard Deviation

❑ Let $E(X) = \mu$

$$Var(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$$

$$sd(X) = \sqrt{Var(X)}$$

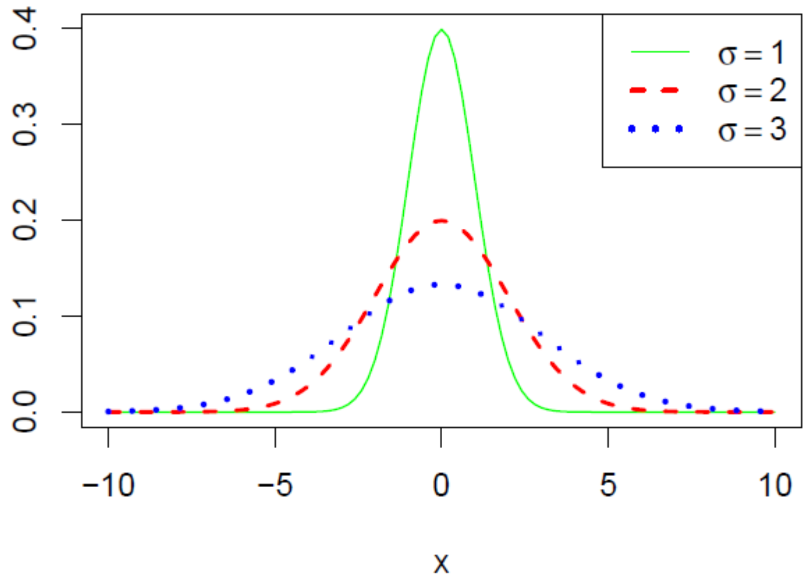


Figure 2: Normal distributions with different Standard deviations

❑ Standard deviation is easier to interpret than variance because it has the same unit as the expected value.

❑ Properties of variance and standard deviation

1. Variance of any constant is zero
2. For any constant a and b, $\mathrm{Var}(aX + b) = a^2 Var(X)$
3. The variance of the sum is NOT equal to the sum of variances.

Standardizing a random variable (Z transform of X)

✓ Different random variables can have very different units!

- Number of heads when flipping a coin 3 times (0,1,2,3)
- Stock price in the past 12 months! ($120-220)
- Height of US teenagers at age 11 (110-160 cm)

✓ These random variables can be transformed (standardized) such that they have the same mean (0) and the same variance (1)

$$Z = \frac{X-\mu}{\sigma}$$

Z measures how many standard deviations X is away from it's mean.

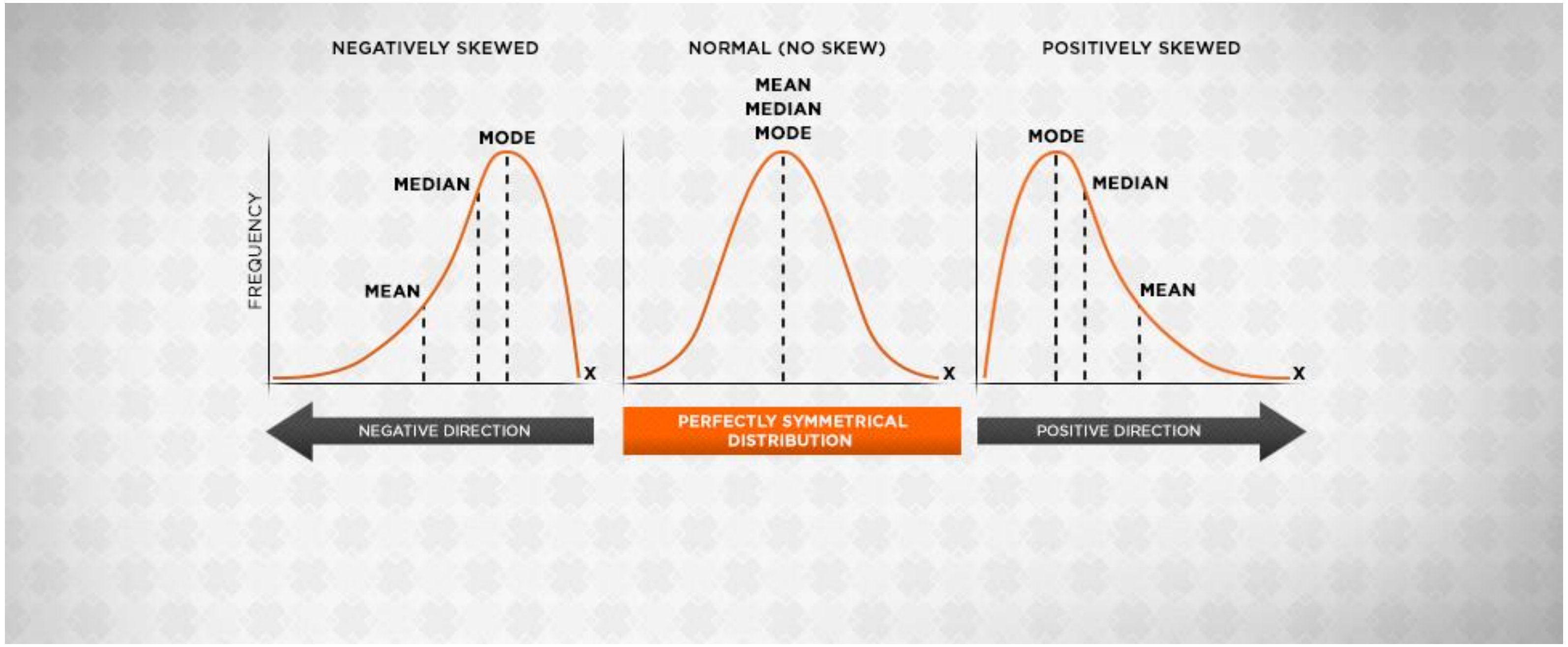
✓ More importantly, the probability distribution of these standardized random variables are often virtually identical!

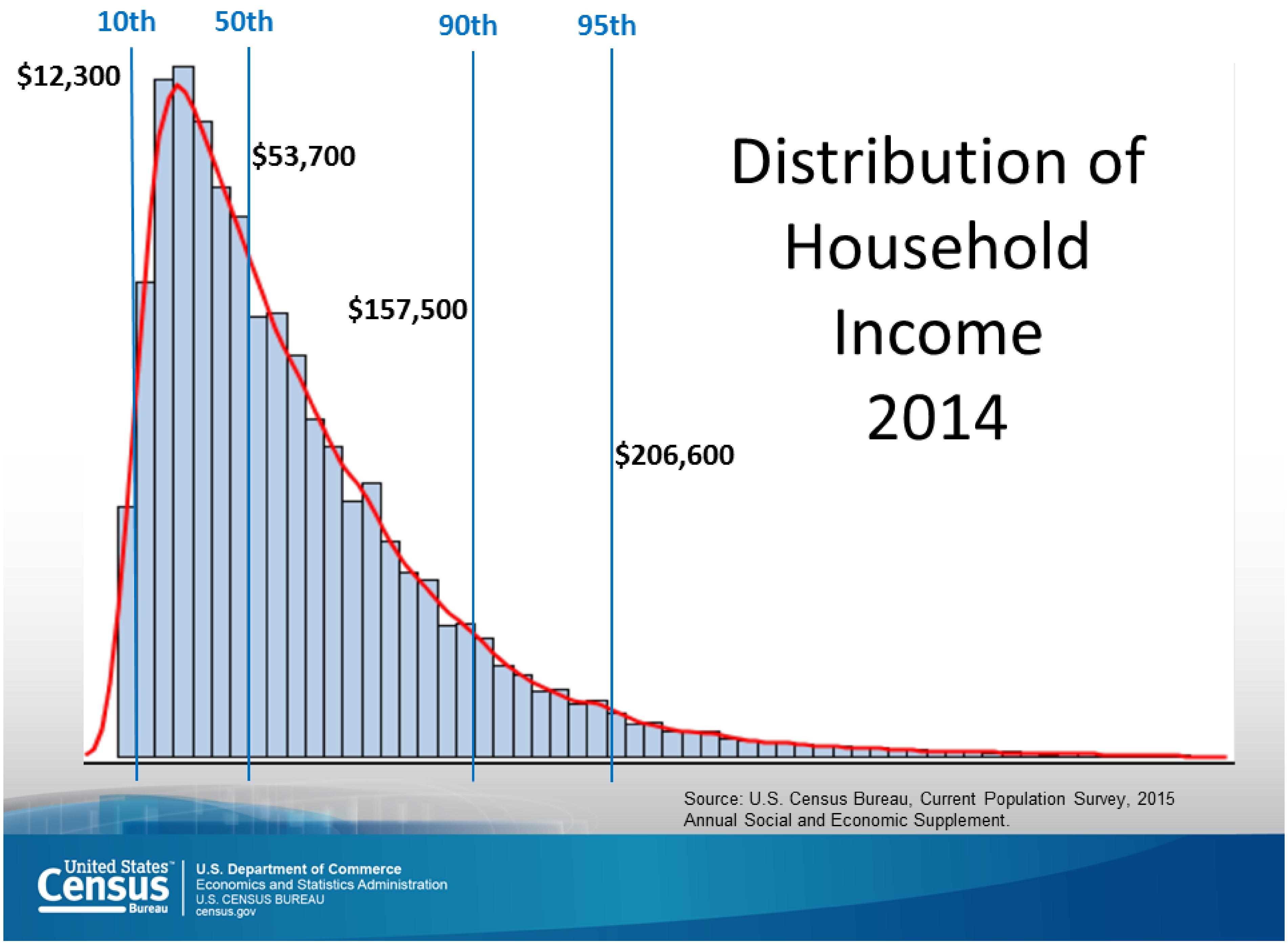
Features of probability distributions

✓ Measures of symmetricity:

1. Skewness: Measures whether a distribution is symmetric about its mean!

$$E(Z^3) = E[(X-\mu)^3]/\sigma^3$$





10th
50th
90th
95th
$12,300
$53,700
$157,500
$206,600
Distribution of Household Income 2014
Source: U.S. Census Bureau, Current Population Survey, 2015 Annual Social and Economic Supplement.
United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Features of joint distributions

✓ Measures of Association (Joint variability):

1. **Covariance**: How, on average, two random variables vary with one another.

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X\mu_Y$$

❑ Properties of Covariance

1. If X and Y are independent, then Cov(X,Y)= 0
2. For any constant a_1, a_2, b_1, b_2 , $Cov(a_1X + b_1, a_2Y + b_2) = a_1a_2Cov(X,Y)$

❑ Variance of sums of random variables

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2ab\ Cov(X,Y)$$

Implication! If X and Y are negatively correlated, then the Var(sum) < Sum(var)

Features of joint distributions

2. Correlation: On average, measures the linear relationship between two random variables

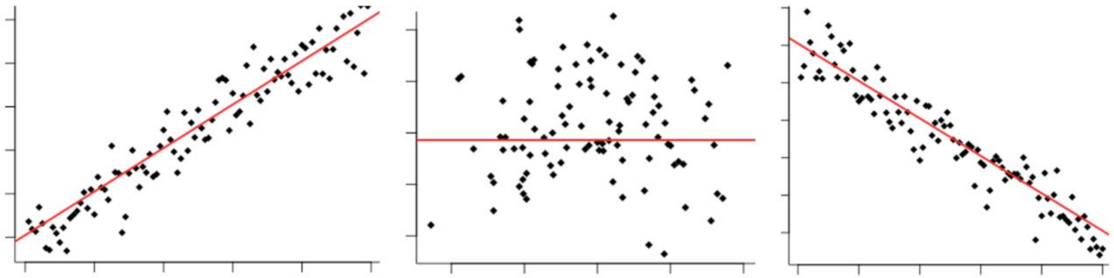
Covariance depends on units! Example: Cov(education, earnings ($ or thousand $))

Correlation fixes this deficiency of covariance:

$$Corr(X,Y) = \frac{Cov(X,Y)}{sd(X)\, sd(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \rho_{XY}$$

❑ Properties of Correlations

1. $-1 \leq Corr(X,Y) \leq 1$



2. For any constant a_1, a_2, b_1, b_2 , $Corr(a_1 X + b_1, a_2 Y + b_2) = Corr(X,Y)$

Independence VS. Mean independence VS. Zero Correlation

Independence $\Rightarrow$ Mean independence $\Longrightarrow$ Zero correlation

$$P(Y|X) = P(Y) \Rightarrow E(Y|X) = E(Y) \Rightarrow Corr(Y,X) = 0$$

BUT

Independence $\nLeftarrow$ Mean independence $\nLeftarrow$ Zero correlation

- ✓ Conditional expectation captures the **nonlinear** relationship between X and Y.
- ✓ Correlation captures the **linear** relationship between X and Y