Econometrics: Lecture 2

Hang Miao

Rutgers University

January 21, 2021

Overview

1. Discrete Random Variables

2. Continuous Random Variables

3. Moments of Distribution

Discrete Random Variables

Definition (Discrete Random Variables)

A discrete random variable X takes on values x_i with probability p_i, $i = 1, \ldots, m$, where $\sum_{i=1}^{m} p_i = 1$.

Definition (Probability Mass Function)

A probability mass function (pmf) is a function that gives the probability that a discrete random variable is exactly equal to some value. The probability mass function is often the primary means of defining a discrete probability distribution.

Definition (Cumulative Distribution Function)

In probability theory and statistics, the cumulative distribution function (CDF) of a real-valued random variable X, or just distribution function of X, evaluated at x, is the probability that X will take a value less than or equal to x.

Roll a Fair Die

Example 1

Roll a fair die and let X be the value that appears. Then X takes on the values 1 through 6, each with probability $1/6$.

R Simulation

```
sample(1:6, 1)
```

Table: PMF and CDF of a Dice Roll

Outcome	1	2	3	4	5	6
PMF	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$
CDF	$1/6$	$2/6$	$3/6$	$4/6$	$5/6$	1

Bernoulli Trials

Example 2

We might as well simulate coin tossing with outcomes H (heads) and T (tails).

R Simulation

```
sample(c("H", "T"), 1)
```

Binomial Distribution

Example 3

It is a well known result that the number of successes k in a Bernoulli experiment follows a binomial distribution. We denote this as

$$k \sim B(n, p).$$

Definition (PMF)

The probability of observing k successes in the experiment $B(n, p)$ is given by

$$f(k) = P(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \frac{n!}{k!(n-k)!} \cdot p^k \cdot (1-p)^{n-k}$$

with $\binom{n}{k}$ the binomial coefficient.

R for Bernoulli and Binomial Distribution

R PMF

```
dbinom(x, size, prob, log = FALSE)
```

R CDF

```
pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)
```

R Quantile Function

```
qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)
```

R Random Generator

```
rbinom(n, size, prob)
```

Continuous Random Variables

No PMF

If a random variable X can take on any of a continuum of values, say, any value between 0 and 1, then we cannot define it by listing values x_i and giving the probability p_i that $X = x_i$; for any single value x_i, $\text{Prob}(X = x_i)$ is zero! Instead we can define the *cumulative distribution function* (CDF) or the *probability density function* (PDF):

Definition (Cumulative Distribution Function)

$$F(x) \equiv \text{Prob}(X < x),$$

PDF is Implicitly Defined

Definition (Probability Density Function)

$$\rho(x)\,dx \equiv \text{Prob}(X \in [x, x + dx]) = F(x + dx) - F(x).$$

Letting $dx \to 0$, we find

$$\rho(x) = F'(x), \quad F(x) = \int_{-\infty}^{x} \rho(t)\,dt.$$

Uniform Distribution in $[0, 1]$

Definition (Uniform Distribution)

The continuous uniform distribution all intervals of the same length on the distribution's support are equally probable.

Definition (CDF)

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

Definition (PDF)

$$\rho(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x > 1 \end{cases}$$

Normal (Gaussian) Distribution, Mean μ, Variance σ^2

Definition (Uniform Distribution)

Normal distributions are important in statistics and are often used in the natural and social sciences to represent real-valued random variables whose distributions are not known

Definition (CDF)

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

Definition (PDF)

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

First Order Moments: Expectation

Definition (Expectation Discrete case)

$$E(X) = \sum_{i=1}^{m} p_i x_i.$$

Definition (Expectation Continuous case)

$$E(X) = \int_{-\infty}^{\infty} x\rho(x)\,\mathrm{d}x.$$

Roll Dice Example

$$E(X) = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 4 + \frac{1}{6}\cdot 5 + \frac{1}{6}\cdot 6 = \frac{7}{2}.$$

Property of Expectation

Important Property

$$E(g(X)) = \sum_{i=1}^{m} p_i g(x_i).$$

Form 1

$g(X) = aX + b$, a and b constants.

$$\begin{aligned} E(g(X)) &= \sum_{i=1}^{m} p_i(ax_i + b) \\ &= a\sum_{i=1}^{m} p_i x_i \ + \ b \quad (\text{since } \sum_{i=1}^{m} p_i = 1) \\ &= a \cdot E(X) + b. \end{aligned}$$

Property of Expectation

Form 2

$g(X) = X^2$. Then $E(g(X)) = \sum_{i=1}^{m} p_i x_i^2$.

Form 2: Roll Dice Example

$$E(X^2) = \frac{1}{6} \cdot 1^2 + \frac{1}{6} \cdot 2^2 + \frac{1}{6} \cdot 3^2 + \frac{1}{6} \cdot 4^2 + \frac{1}{6} \cdot 5^2 + \frac{1}{6} \cdot 6^2 = \frac{91}{6}.$$

Second Order Moments: Variance

Definition (Variance)

Let $\mu = E(X)$ denote the expected value of X. Then the Variance is defined as the expected value of the *square of the difference* between X and μ, i.e $Var(X) = E((X-\mu)^2)$

Equivalent Representation

$$\begin{aligned} E((X-\mu)^2) &= \sum_{i=1}^{m} p_i(x_i-\mu)^2 \\ &= \sum_{i=1}^{m} p_i(x_i^2 - 2\mu x_i + \mu^2) \\ &= \sum_{i=1}^{m} p_i x_i^2 - 2\mu \sum_{i=1}^{m} p_i x_i + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

Property of Variance

Important Property

Suppose X is a random variable. Y is the linear transformation of of X such that $Y = a \cdot X + b$. Then

$$Var[Y] = a^2 Var[X]$$

$$\begin{aligned}
Var[Y] &= E[(Y - E[Y])^2] \\
&= E[(a \cdot X + b - E[a \cdot X + b])^2] \\
&= E[(a \cdot X + b - a \cdot E[X] - b)^2] \\
&= E[a^2 \cdot (X - E[X])^2] \\
&= a^2 \cdot E[(X - E[X])^2] \\
&= a^2 \cdot Var[X]
\end{aligned}$$

Statistics Involved with Third Order Moments: Skewness

Skewness

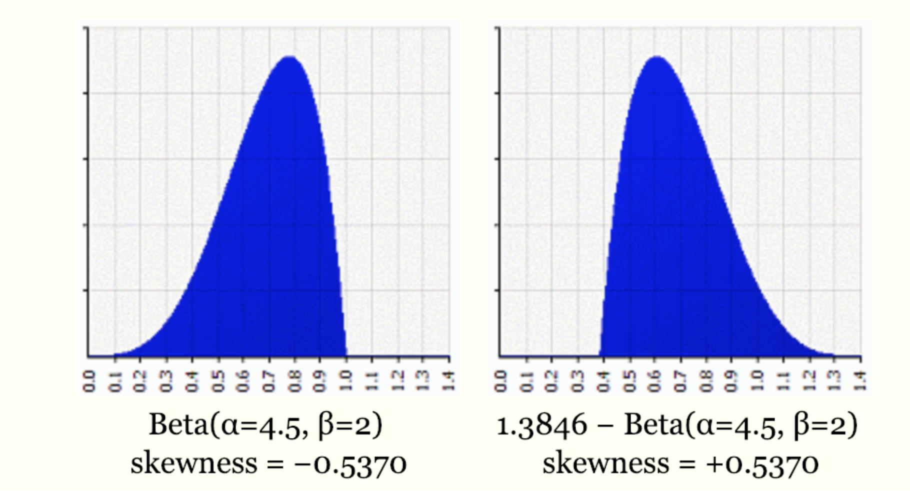
For a unimodal distribution, skewness is a measure of the asymmetry of that probability distribution about its mean. The skewness value can be positive or negative, or undefined (undefined if the distribution is not unimodal, i.e has more than one peak).

Definition (Pearson's skewness)

There many different numerical ways to measure skewness. Pearson's skewness is one of the most commonly used one.

$$
\begin{aligned}
Skew[X] &= \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \\
&= \frac{\mathrm{E}[X^3] - 3\mu\,\mathrm{E}[X^2] + 3\mu^2\,\mathrm{E}[X] - \mu^3}{\sigma^3} \\
&= \frac{\mathrm{E}[X^3] - 3\mu\sigma^2 - \mu^3}{\sigma^3}
\end{aligned}
$$

Examples of Skewness



Properties of Skewness

Sign of Skewness

- If skewness is positive, the data are positively skewed or skewed right, meaning that the right tail of the distribution is longer than the left.
- If skewness is negative, the data are negatively skewed or skewed left, meaning that the left tail is longer.
- If skewness $= 0$, the data are perfectly symmetrical.

Bulmer (1979)'s rule of thumb:

But a skewness of exactly zero is quite unlikely for real-world data

- If skewness is less than -1 or greater than $+1$, the distribution is highly skewed.
- If skewness is between -1 and $-\frac{1}{2}$ or between $+\frac{1}{2}$ and $+1$, the distribution is moderately skewed.
- If skewness is between $-\frac{1}{2}$ and $+\frac{1}{2}$, the distribution is approximately symmetric.

Statistics Involved with Fourth Order Moments: Kurtosis

Kurtosis

Kurtosis is a measure of the "tailedness" of the probability distribution. Higher kurtosis means more of the variance is the result of infrequent extreme deviations, as opposed to frequent modestly sized deviations.

Definition (Pearson's Kurtosis)

There many different numerical ways to measure Kurtosis. Pearson's skewness is one of the most commonly used one.

$$\mathrm{Kurt}[X] = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$$

Definition (Excess Kurtosis)

The reference standard is a normal distribution, which has a kurtosis of 3. In token of this, often the excess kurtosis is presented: excess kurtosis is simply $\mathrm{Kurt} - 3$

Properties of Kurtosis

Mesokurtic

Distributions with zero excess kurtosis are called mesokurtic. The most prominent example of a mesokurtic distribution is the normal distribution family, regardless of the values of its parameters.

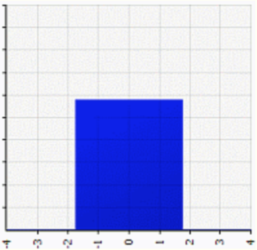
Leptokurtic

A distribution with positive excess kurtosis is called leptokurtic, or leptokurtotic. "Lepto-" means "slender".

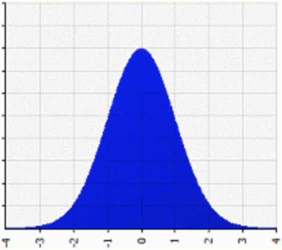
Platykurtic

A distribution with negative excess kurtosis is called platykurtic, or platykurtotic. "Platy-" means "broad". In terms of shape, a platykurtic distribution has thinner tails. Examples of platykurtic distributions include the continuous and discrete uniform distributions,

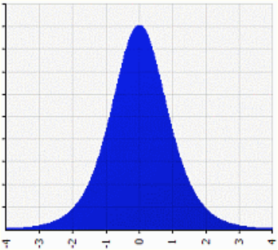
Examples of Kurtosis



Uniform(min=$-\sqrt{3}$, max=$\sqrt{3}$)
kurtosis = 1.8, excess = −1.2

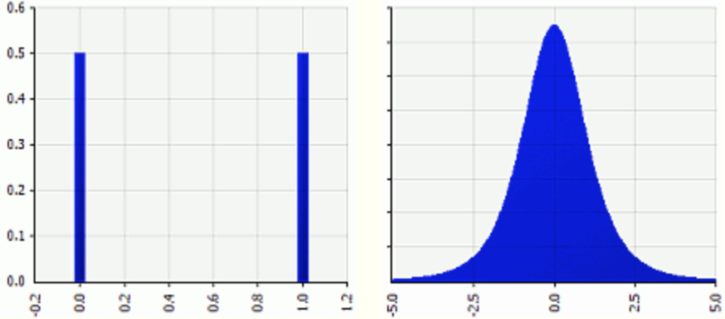


Normal(μ=0, σ=1)
kurtosis = 3, excess = 0



Logistic(α=0, β=0.55153)
kurtosis = 4.2, excess = 1.2

Examples of Kurtosis



Discrete: equally likely values
kurtosis = 1, excess = −2

Student's t (df=4)
kurtosis = ∞, excess = ∞