ECON 3740: INTRODUCTION TO ECONOMETRICS

INSTRUCTOR: CHAOYI CHEN
Department of Economics and Finance, University of Guelph

Lecture 2

Lecture outline

Last lecture, we reviewed the definition of a random variable (both discrete and continuous) and its probability distribution. Today, we will go through

- Measures of the shape of a probability distribution (mean, variance)
- Two random variables and their joint distribution
  - Joint distribution, marginal distribution, conditional distribution
  - Law of iterated expectations
  - Means, variances and covariances of sums of random variables
- Often used probability distributions in econometrics
  - Normal, Chi-Squared, Student t and F-distributions

- The *expected value* or mean of a random variable is the average value over many repeated trails or occurrences.

Definition: Expected value of a discrete random variable Y with k possible values

$$E(Y) = \sum_{i=1}^{k} y_i P(Y = y_i) = \mu_y$$

- Let random variable G denote number of days it will snow in the next week of January

Probability distribution of G								
Outcome	0	1	2	3	4	5	6	7
Probability	0.20	0.25	0.20	0.15	0.10	0.05	0.04	0.01

Question: what is the expected value for the random variable G?

Expected value for a continuous random variable

Definition: Expected value of a continuous random variable Y (if the domain for y is (-∞, ∞) and the PDF for y is $f(y)$)

$$E(Y) = \int_{-\infty}^{\infty} yf(y)dy = \mu_y$$

- Let random variable T denote next monday temperature. Suppose the domain for T is $[-10, 10]$.

Question: Assuming that T satisfy the uniform distribution with probability density $f(t) = \frac{1}{20}$, what is the expected value for the random variable T?

The variance for a discrete random variable

- The *The variance* of a random variable Y is the expected value of the square of the deviation of Y from its mean.

Definition: The variance of a discrete random variable Y with k possible values

$$Var(Y) = E[(Y - \mu_y)^2] = \sum_{i=1}^{k} (y_i - \mu_y)^2 P(Y = y_i) = \sigma_y^2$$

- Let random variable T denote number of days it will snow in the next week of January

Probability distribution of G								
Outcome	0	1	2	3	4	5	6	7
Probability	0.20	0.25	0.20	0.15	0.10	0.05	0.04	0.01

Question: what is the variance for the random variable G?

The variance for a continuous random variable

Definition: Variance of a continuous random variable Y (if the domain for y is (-∞, ∞) and the PDF for y is $f(y)$)

$$Var(Y) = E[(Y - \mu_y)^2] = \int_{-\infty}^{\infty} (y - \mu_y)^2 f(y) dy = \sigma_y^2$$

- Let random variable T denote next monday temperature. Suppose the domain for T is $[-10, 10]$.

Question: Assuming that T satisfy the uniform distribution with probability density $f(t) = \frac{1}{20}$, what is the expected value for the random variable T?

Mean and variance of a Bernoulli random variable

- *A Bernoulli random variable* is a binary random variable with two possible outcomes, 0 and 1.
- For instance, let B be a random variable which equals 1 if you pass the exam and 0 if you dont pass

$$B = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } (1-p) \end{cases}$$

Question: what is the expected value of B and the variance of B?

Mean and variance of a linear function of a random variable

Consider two random variables (X, Y) that are related by a linear function

$$Y = a + bX$$

- Assume that $E(X) = \mu_X$ and $Var(X) = \sigma_X^2$

Question: what is the mean and variance for Y?

Two random variables and their joint distribution

- Many intriguing problems in economics involve 2 or more random variables
- We need to understand the concepts of joint, marginal and conditional probability distribution to solve those problems

Definition: The joint probability distribution of two discrete random variables X, Y is

$$P(X = x \textit{ and } Y = y)$$

An example:

- Let Y equal 1 if it rains and 0 if it does not rain.
- Let X equal 1 if it is humid and 0 if it is not humid.

Joint probability distribution of X and Y

	humid ($X = 1$)	No humid ($X = 0$)	Total
Rain (Y=1)	0.15	0.07	0.22
No Rain (Y=0)	0.15	0.63	0.78
Total	0.3	0.7	1

Two random variables and the marginal distributions

The conditional distribution is the distribution of a random variable conditional on another random variable taking on a specific value.

- The conditional probability that it rains given that it is humid

$$P(Y=1|X=1)=\frac{P(Y=1,X=1)}{P(X=1)}=\frac{0.15}{0.3}=0.5$$

- In general, the conditional probability (distribution) of Y given X is

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

- The conditional expectation of Y given X is

$$E(Y|X=x)=\sum_{i=1}^{k} y_i P(Y=y_i|X=x)$$

Question: what is the expected value of rain given that it is humid?

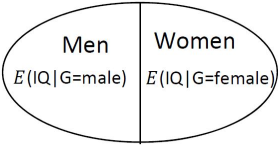
The law of iterated expectations

Law of iterated expectations states that the mean of Y is the weighted average of the conditional expectation of Y given X, weighted by the probability distribution of X.

$$E(Y) = E[E(Y|X)] = \sum_{i=1}^{k} E(Y|X = x_i)P(X = x_i)$$

An example:

- Suppose we are interested in average IQ generally, but we have measures of average IQ by gender.



$$E(IQ) = E[E(IQ|G)] = E(IQ|G = m)P(G = m) + E(IQ|G = f)P(G = f)$$

Covariance

- The *covariance* is a measure of the extend to which two random variables X and Y move together

$$\text{Cov(X,Y)=E[(X-}_m u_X)(Y = \mu_Y)]$$
$$=\sum_{i=1}^{k}\sum_{j=1}^{l}(x_j - \mu_X)(y_i - \mu_Y)P(X = x_j, Y = y_i)$$

- Recall the humidity and rain example

Joint probability distribution of X and Y			
	humid ($X = 1$)	No humid ($X = 0$)	Total
Rain (Y=1)	0.15	0.07	0.22
No Rain (Y=0)	0.15	0.63	0.78
Total	0.3	0.7	1

Question: what is the covariance between rain (Y) and humid (X)?

Correlation

- Why we need correlation? it is units free. Recall that the covariance of X and Y are their units multiplication, which, sometimes, is hard to interpret the size
- The correlation between X and Y is defined as

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

Properties:

- A correlation is always between -1 and 1 and X and Y are uncorrelated if $Corr(X, Y) = 0$
- If the conditional mean of Y does not depend on X, X and Y are uncorrelated

$$if \;\; E(Y|X) = E(Y), then \; Cov(X, Y) = 0 \;\&\; Corr(X, Y) = 0$$

Question: if X and Y are uncorrelated, does this necessarily imply they are independent?

Means, Variances and covariances of sums of random variables

- Let G=aX+bY
- Given the $E(X) = \mu_X$, $E(Y) = \mu_Y$, what is the mean of G?
- The variance of G is

$$Var(G) = Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$$

Please prove above result at home as a practice question.

Some useful results

Means, Variances, and Covariances of Sums of Random Variables

KEY CONCEPT 2.3

Let X, Y, and V be random variables, let μ_X and σ_X^2 be the mean and variance of X, let σ_{XY} be the covariance between X and Y (and so forth for the other variables), and let a, b, and c be constants. Equations (2.29) through (2.35) follow from the definitions of the mean, variance, and covariance:

$$E(a + bX + cY) = a + b\mu_X + c\mu_Y, \quad (2.29)$$

$$\mathrm{var}(a + bY) = b^2\sigma_Y^2, \quad (2.30)$$

$$\mathrm{var}(aX + bY) = a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2, \quad (2.31)$$

$$E(Y^2) = \sigma_Y^2 + \mu_Y^2, \quad (2.32)$$

$$\mathrm{cov}(a + bX + cV, Y) = b\sigma_{XY} + c\sigma_{VY}, \quad (2.33)$$

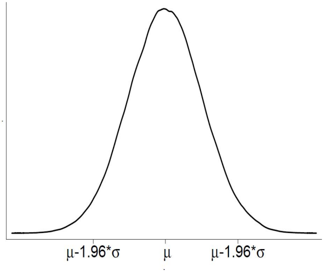
$$E(XY) = \sigma_{XY} + \mu_X\mu_Y, \quad (2.34)$$

$$|\mathrm{corr}(X, Y)| \le 1 \text{ and } |\sigma_{XY}| \le \sqrt{\sigma_X^2\sigma_Y^2} \text{ (correlation inequality).} \quad (2.35)$$

Often used probability distributions in Econometrics - normal distribution

- The most often used probability density function in econometrics is the Normal (or Gaussian) distribution. The normal distribution is useful because of the central limit theorem. If a random variable $Y \sim N(\mu, \sigma^2)$, then the density function for Y is

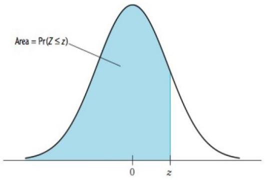
$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{(y-\mu)}{\sigma})^2}$$



Often used probability distributions in Econometrics - normal distribution - continue

- A standard normal distribution $N(0,1)$ has mean $= 0$ and variance $= 1$
- A random variable with a $N(0,1)$ distribution is denoted by Z and its CDF, $\phi(z) = P(Z \leq z)$, can be found in Z table

TABLE 1 The Cumulative Standard Normal Distribution Function, $\Phi(z) = \Pr(Z \leq z)$



Second Decimal Value of z

z	0	1	2	3	4	5	6	7	8	9
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048

Often used probability distributions in Econometrics - normal distribution - continue

- Consider a random variable Y, which is normally distributed with $mean = \mu$ and variance$= \sigma^2$
- To calculate the probability, we must first *standardize* Y to get the standard normal variable Z. That is

$$Z = \frac{(Y - \mu)}{\sigma}$$

- Then, you can check Z table to look up the probabilities

An example

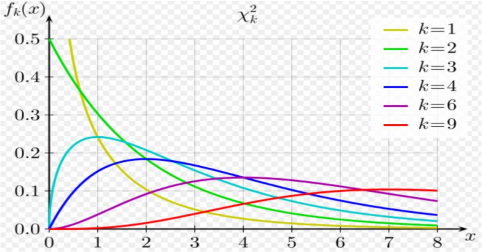
- Let $Y \sim N(5, 2)$

$$\text{P}(\text{Y} \leq 0) = P\left(\frac{(Y-5)}{2} \leq \frac{(0-5)}{2}\right)$$
$$= \text{P}\left(\text{Z} \leq -2.5\right) = 0.0062$$

Often used probability distributions in Econometrics - chi-square distribution

- The *chi-squared* distribution is the distribution of the **sum** of m squared independent **standard** normal random variables
- Specifically, let $Z_1, Z_2, .., Z_k$ be k independent standard normal random variables. Then, the sum of the squares of these random variables forms a chi-squared distribution with k degrees of freedom

$$\sum_{i=1}^{k} Z_i^2 \sim \chi_k^2$$



Often used probability distributions in Econometrics - chi-square distribution - continue

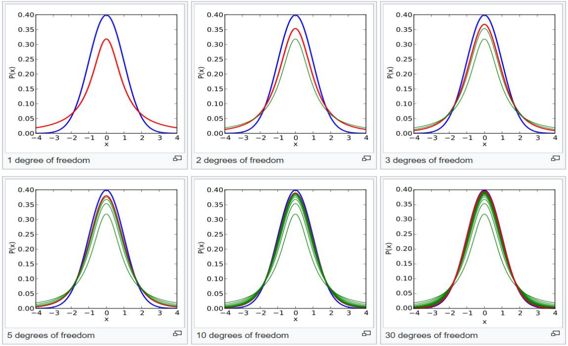
- To look up probabilities of a chi-square distribution, you should check chi-square table
- For example P($\sum_{i=1}^{3} Z_i^2 \leq 7.81$)= ?

Chi-square Distribution Table

d.f.	.995	.99	.975	.95	.9	.1	.05	.025	.01
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21

Often used probability distributions in Econometrics - student t distribution

- Let Z be a standard normal random variable and W be a chi-squared distributed random variable with k degree of freedom
- The student t distribution with k degrees of freedom is the distribution with random variable $T = \frac{Z}{\sqrt{W/k}}$



Often used probability distributions in Econometrics - student t distribution-continue

- The Student t distribution is often used when testing hypotheses in econometrics
- You can check T table to find the CDF. For example with $k = 5$

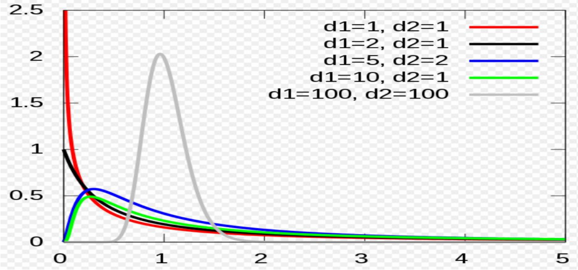
$$\text{P}(\text{t} \leq -2.57) = 0.025, P(t \geq 2.57) = 0.025$$

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869

Often used probability distributions in Econometrics - F distribution

- Let W a chi-squared random variable with k degrees of freedom and V a chi-squared random variable with n degrees of freedom.
- The F-distribution with k and n degrees of freedom Fm;n is the distribution of the random variable $F = \frac{W/k}{V/n}$



Often used probability distributions in Econometrics - F distribution - continue

- Check F table to find the probability
- For example, with $k = 4$, $n = 1$,

$$\text{P}(\text{F} \geq 4.54) = 0.1, P(F \geq 7.71) = 0.05$$

		Degrees of freedom in the numerator								
om in the denominator	p	1	2	3	4	5	6	7	8	9
1	.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
	.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28
	.010	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
	.001	405284	500000	540379	562500	576405	585937	592873	598144	602284
2	.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
	.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
	.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39
3	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86
4	.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66