Quantitative Economics Lecture 4 - Multivariate Distributions

Richard Povey

University of Oxford

30th April 2018

richard.povey@hertford.ox.ac.uk, richard.povey@st-hildas.ox.ac.uk

With thanks to Ian Crawford, Margaret Stevens, Vanessa Berenguer-Rico.

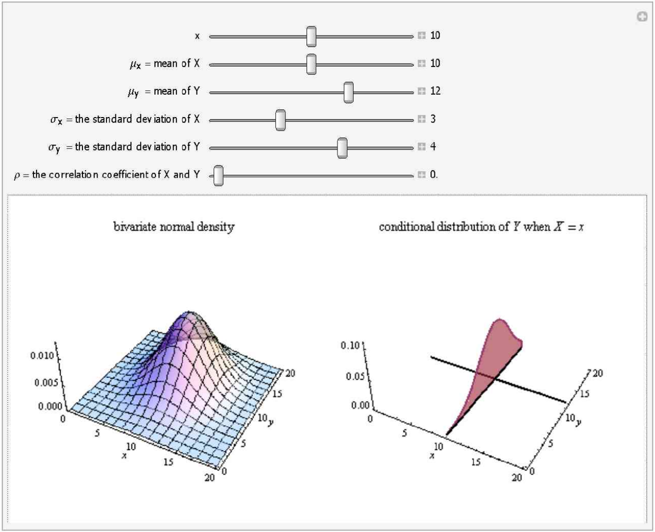
Outline of Lecture

- In this lecture we will look at some important multivariate distributions which are derived from or represent extensions of the standard normal distribution:
- **Bivariate normal** - Useful for visualising correlation, and for revisiting the concept of correlation in more detail and introducing the concept of a **conditional expectation**.
- **Chi-Square variables** - sum of the squares of independent normally distributed variables - important for testing *multiple* hypotheses simultaneously.
- **Student's t and Snedecor-Fisher F distributions** - Important for statistical inference, but we will see that for most applications in QE with large sample size they are asymptotically equivalent to standard normal or chi-square distributions.

Bivariate Normal Distribution

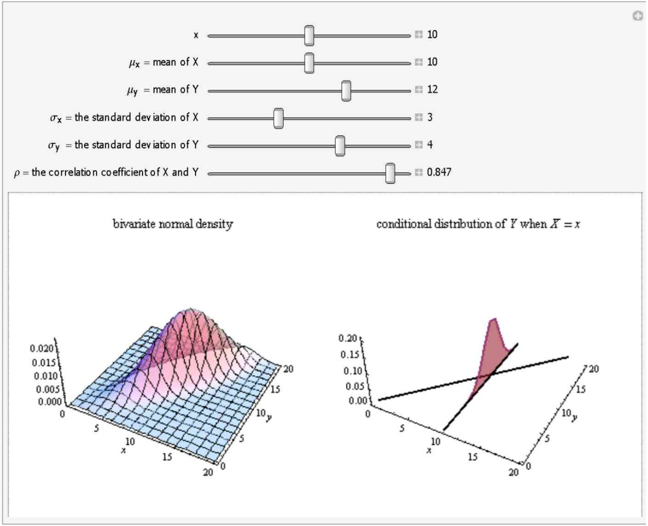
Note: The line in the right hand diagram shows $E[Y|X]$.

From Wolfram Demonstrations Project



Note: The line in the right hand diagram shows $E[Y|X]$.

From Wolfram Demonstrations Project



Definition

Conditional Expectation (Bivariate case): Let Y have finite expectation. The conditional expectation of Y given $X = x$ is defined as the mean of the conditional density of Y given $X = x$.

- (i) Discrete case:

$$E[Y|X = x_i] = \sum_{j=1}^{l} \left[y_j f_{Y|X}(y_j|X = x_i)\right] = \sum_{j=1}^{l} \left[y_j \left(\frac{f_{X,Y}(x_i, y_j)}{f_X(x_i)}\right)\right]$$

- (ii) Continuous case:

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|X = x)\, dy = \int_{-\infty}^{\infty} y \left(\frac{f_{X,Y}(x, y)}{f_X(x)}\right) dy$$

Conditional Expectation Properties

- Note: In statistics $m(x) = E[Y|X = x]$ is called the **regression function**.

Properties of $E[Y|X = x]$: Let $E[|Y|] < \infty$. Then:

- (i) $E[a + bY|X = x] = a + bE[Y|X = x]$
- (ii) **Law of Iterated Expectations**: $E[E[Y|X = x]] = E[Y]$
- (iii) If Y and X are **independent**, then $E[Y|X = x] = E[Y]$

We assume X and Y are continuously distributed random variables:

$$\text{(i)}\ E\left[a + bY|X = x\right] = \int_{-\infty}^{\infty} (a + by) f_{Y|X}\left(y|X = x\right) dy$$

$$= a \int_{-\infty}^{\infty} f_{Y|X}\left(y|X = x\right) dy + b \int_{-\infty}^{\infty} y f_{Y|X}\left(y|X = x\right) dy$$

$$= a + bE[Y|X = x]$$

$$\text{(ii)}\ E\left[E\left[Y|X = x\right]\right] = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y f_{Y|X}\left(y|X = x\right) dy \right) f_X(x) dx$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y \left(\frac{f_{X,Y}\left(x, y\right)}{f_X(x)} \right) dy \right) f_X(x) dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{X,Y}\left(x, y\right) dy\ dx$$

$$= E[Y]$$

Conditional Expectation Properties

Properties of $E[Y|X=x]$: Let $E[|Y|] < \infty$. Then:

- (iv) If $u = Y - E[Y|X=x]$, then $E[g(X)u] = 0$ for any function $g(X)$ provided $E(|g(X)u|) < \infty$. In particular, $E(u) = 0$ and $E(Xu) = 0$

- (v) **Conditional Jensen's Inequality**: If $g(X)$ is a convex function , then $g(E[Y|X=x]) \leq E[g(Y)|X=x]$

- (vi) If $E(Y^2) < \infty$, then $E[Y|X=x]$ is a solution to $\min_{m \in M} \left\{ E\left[(Y - m(X))^2 \right] \right\}$ where M is the set of possible solutions. That is: $E[Y|X=x]$ is the **best mean square predictor** of Y given X.

Correlation Revisited

$$Corr(Y,X) = \frac{Cov(Y,X)}{\sqrt{Var(Y)Var(X)}} = \rho_{XY} \text{ where } -1 \leq \rho_{XY} \leq 1$$

- Correlation between two random Variables X and Y is:
  - Positive when X and Y increase and decrease together: if X is high (low), then there is a high probability that Y is high (low).
  - Negative with high values of X are associated with low values of Y and vice versa
  - Zero when they are independent
- **Important**: Uncorrelated does *not* imply independence. It is possible for correlation to be zero even when X and Y are related - even *perfectly* related as in the example below - if the relationship is non-linear:

Corr(X,Y) = 1

	X=1	X=2	X=3
Y=1	0.111	0	0
Y=2	0	0.4445	0
Y=3	0	0	0.4445

Corr(X,Y) = -1

	X=1	X=2	X=3
Y=1	0	0	0.4445
Y=2	0	0.4445	0
Y=3	0.111	0	0

Corr(X,Y) = 0 but X and Y are NOT Independent

	X=1	X=2	X=3
Y=1	0.111	0	0
Y=2	0	0	0.4445
Y=3	0	0.4445	0

Correlation and Law of Iterated Expectations Example

Example:

$$\begin{aligned}
\mathrm{Cov}(W,H) &= (1-0.2)(1-0.14)\times 0.09 + \\
&\quad (0-0.2)(1-0.14)\times 0.05 + \\
&\quad (1-0.2)(0-0.14)\times 0.11 + \\
&\quad (0-0.2)(0-0.14)\times 0.75 \\
&= 0.062 \\
\mathrm{Corr}(W,H) &= 0.45
\end{aligned}$$

Joint Distribution of W and H			
	W=1	W=0	Total
H=1	0.09	0.11	0.20
H=0	0.05	0.75	0.80
	0.14	0.86	1.00
Conditional Distribution of W given H			
P(W|H=1)	0.45	0.55	1.00
P(W|H=0)	0.06	0.94	1.00

$$\mathrm{E}(W) = 0.14$$
$$\mathrm{E}(H) = 0.20$$
$$\mathrm{Var}(W) = 0.14 \times 0.86 = 0.1204$$
$$\mathrm{Var}(H) = 0.20 \times 0.80 = 0.16$$

Law of Iterated Expectations - Discrete Example

$$E[W] = E\left[E[W|H]\right] = \sum_{i=1}^{k}\left[\sum_{j=1}^{l}\left[w_j f_{W|H}\left(W = w_j | H = h_i\right)\right] f_H(h_i)\right]$$
$$= \sum_{i=1}^{2}\left[E[W|H=h_i] f_H(h_i)\right] = f_H(1)E[W|H=1] + f_H(0)E[W|H=0]$$
$$= 0.2\times(1\times 0.45 + 0\times 0.55) + 0.8\times(1\times 0.0625 + 0\times 0.9375) \; = \underline{0.14}$$

Independence

Independence and correlatedness

$$X \text{ and } Y \text{ are independent} \Rightarrow Cov(X, Y) = 0$$

$$X \text{ and } Y \text{ are independent} \Rightarrow Corr(X, Y) = 0$$

The converse, that if $Cov(X, Y) = 0$ or $Corr(X, Y) = 0$ they must be independent, is not true.

Mean-independence

- **Mean-independence** is a notion of independence which lies "between" independence and uncorrelatedness. It plays an important role in regression analysis.

$$E[Y|X] = E[Y] \iff Y \text{ is mean-independent of } X$$

$$E[X|Y] = E[X] \iff X \text{ is mean-independent of } Y$$

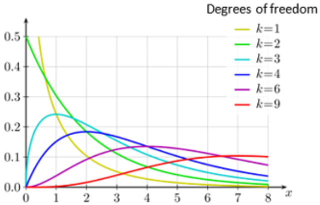
$$\text{Independence} \Rightarrow \text{Mean-independence}$$

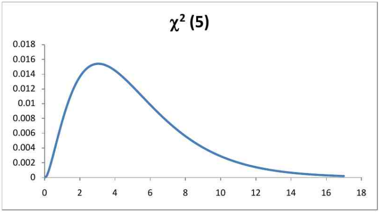
$$\text{Independence} \not\Leftarrow \text{Mean-independence}$$

- Let $X \sim N(0,1)$ Then $X^2 \sim \chi^2(1)$ for $i = 1, 2, ..., k$. Let $X_i \overset{iid}{\sim} N(0,1)$. Then:

$$W_k = \sum_{i=1}^{k} [X_i^2] \overset{d}{=} \chi^2(k)$$

- The $\chi^2(k)$ distribution is positively-skewed but, by the CLT, as $k \longrightarrow \infty$, $W_k \xrightarrow{d} N(k, 2k)$ so the skewness disappears asymptotically. *(There are exercises in the problem sets asking you to prove that $E[W_k] = k$ and $Var[W_k] = 2k$.)*





Source: https://en.wikipedia.org/wiki/Chi-squared_distribution

Student's t-distribution (William Sealy Gosset)

- Let $Y \sim N(0,1)$ and $X \sim \chi^2(k)$. Then,

$$t = \frac{Y}{\sqrt{\frac{X}{k}}} \stackrel{d}{=} t(k)$$

has a Student's t-distribution with k degrees of freedom.

- The t-distribution is important if (as is usually the case - a lot more on this in Lecture 5) we must estimate the standard deviation of the sample mean from the sample itself.
- For example, suppose we construct our standardised mean using the unbiased and consistent estimator for the population variance:

$$\bar{S}^2 = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left[\left(X_i - \bar{X}\right)^2\right]$$

Student's t-Distribution

- Instead of a z-statistic we then have a t-statistic:

$$t_{\bar{X}} = \frac{\bar{X} - \mu_X}{\sqrt{\frac{\bar{S}^2}{n}}}$$

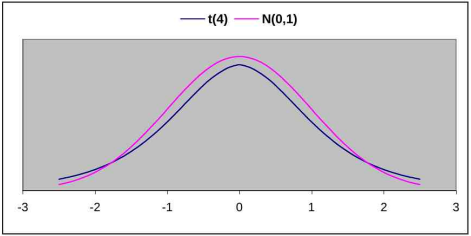
- If the X_is are *precisely* normally distributed, so that $X_i \overset{iid}{\sim} N(\mu_X, \sigma_X^2)$, then $t_{\bar{X}} \overset{d}{=} t(n-1)$.

- Here $\sqrt{\frac{\bar{S}^2}{n}} = \frac{\bar{S}}{\sqrt{n}} = se\left(\bar{X}\right)$ is the standard error of the sample mean $\bar{X}$.

Student's t-Distribution

$$t_{\bar{X}} = \frac{\bar{X} - \mu_X}{\sqrt{\frac{\bar{S}^2}{n}}}$$

- The randomness in the denominator $\sqrt{\frac{\bar{S}^2}{n}}$ due to the sample variation in $\bar{S}^2$ causes *excess kurtosis* in the t-distribution, as can be seen in the graph below comparing the PDFs for the $N(0,1)$ and the $t(4)$ distributions:



Student's t-Distribution

$$t_{\bar{X}} = \frac{\bar{X} - \mu_X}{\sqrt{\frac{\bar{S}^2}{n}}}$$

- However, we often *cannot* assume that the X_is are precisely normally distributed, so what is *much* more important for our purposes is the *asymptotic* distribution of the t-statistic.
- By the Law of Large Numbers, $\sqrt{\frac{\bar{S}^2}{n}} \xrightarrow{p} \sqrt{\frac{\sigma_X^2}{n}}$ (i.e. the standard error we calculate from the sample is a consistent estimator for the asymptotic standard error $\frac{\sigma_X}{\sqrt{n}}$).
- So, by the Central Limit Theorem $t_{\bar{X}} \xrightarrow{d} N(0, 1)$.
- **Important**: This is the key result we employ in Lecture 5 for hypothesis testing.

F-distribution (Snedecor-Fisher)

- Let $Y \sim \chi^2(k)$ be independent of $X \sim \chi^2(l)$ Then,

$$f = \frac{\left(\frac{Y}{k}\right)}{\left(\frac{X}{l}\right)}$$

  has a Snedecor-Fisher F-distribution with (k, l) degrees of freedom:

$$f \stackrel{d}{=} F(k, l)$$

- More detail on the F-distribution and its applications will be given in Lecture 6 on multiple hypothesis testing.