Methods of Mathematical Statistics

Notes by Tim Brown and Guoqi Qian

Module 4: Bivariate Distributions

Contents

1 Generalities, Joint and Marginal Distributions - 4.1 and 4.3

1.1 Review Univariate - Discrete and Continuous Distributions for rv X

Where have we come from?

PMF/PDF For any real x the *Probability Mass/Density Function* $f(x)$, is given by

$$f(x) = \begin{cases} P(X = x), & \text{X discrete} \\ F'(x), & \text{X continuous, F the CDF} \end{cases} \tag{1}$$

Properties (a) $0 \le f(x)$(and ≤ 1 for X discrete) and $f(x) > 0$ on the possible values of X

(b)

$$\begin{cases} \sum_{x \in range(X)} f(x) = 1, & \text{X discrete} \\ \int_{-\infty}^{\infty} f(x)\,dx = 1, & \text{X continuous} \end{cases}$$

(c) for any subset A of reals

$$P(X \in A) = \begin{cases} \sum_{x \in A} f(x), & \text{X discrete} \\ \int_A f(x)\, dx, & \text{X continuous} \end{cases}$$

Where are we going?

Consider two random variables X, Y rather than one.

Assume both are discrete or both are continuous.

Base our calculations to do with X, Y together on a *joint* PMF/PDF.

Aim: Extend our capability to deal with *independent* RVs (Variance, MGF) to *dependent* RVs.

Need to consider *joint* probabilities for X, Y.

1.2 Bivariate - two random variables

New Notation for two random variables X, Y

Write for any sets A, B of real numbers the event $[X \in A] \cap [Y \in B]$ as $[X \in A, Y \in B]$

Example $[X = 1] \cap [Y = 2]$ is written as $[X = 1, Y = 2]$

Example $[X \leq 2] \cap [Y \geq 3]$ is written as $[X \leq 2, Y \geq 3]$

PDF/PDF for rv's X, Y

Analogues of (a) to (c) specify all probabilities for X, Y. PMF/PDF The *Joint Probablity Mass Function/Probability Density Function* $f(x, y)$ satisfies

(a) $0 \leq f(x, y)$(and ≤ 1 for X, Y discrete) and $f(x, y) > 0$ on the *support*, $S_{(X,Y)}$, of the random variables X, Y

(b)

$$\begin{cases} \sum f(x, y) = 1, & \text{X and Y discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx dy = 1, & \text{X and Y continuous} \end{cases}$$

(c) for any subset A of the plane

$$P((X, Y) \in A) = \begin{cases} \sum_{(x,y) \in A} f(x, y), & \text{X and Y discrete} \\ \int \int_A f(x, y)\, dx dy, & \text{X and Y continuous} \end{cases}$$

Marginal PMF/PDF

The PMF or PDF of either random variable is called the *marginal* PMF or PDF because it can be obtained by summing or integrating the joint PMF or PDF over the other variable. For a real subset A

$$P(X \in A) = \begin{cases} \sum_{(x,y)\in A\times\mathbb{R}} f(x,y), & \text{X and Y discrete} \\ \int\int_{A\times\mathbb{R}} f(x,y)\,dxdy, & \text{X and Y continuous} \end{cases}$$

So

$$f_X(x) = \begin{cases} \sum_{y\in\mathbb{R}} f(x,y), & \text{X and Y discrete} \\ \int_{\mathbb{R}} f(x,y)\,dy, & \text{X and Y continuous} \end{cases}$$

Independence of X, Y occurs when the joint PMF or PDF is a product of the marginal pmf's or pdf's.

1.3 Discrete Joint PMF - Accident Category Example

Example - Serious and Non-Serious Accidents

Suppose accidents occur according to a Poisson process of rate λ and that each accident is serious with probability $p > 0$, independent of the process of accident occurrence and of other accidents. Find the joint and marginal pmf's of the number of serious and non-serious accidents in a time interval of 1. Are these two random variables independent?

Solution - Serious and Non-Serious Accidents

Let X be the number of serious accidents in the unit time interval and Y be the number of non-serious accidents. Then for any $x, y \in \{0, 1, 2, \cdots\}$

$$\begin{aligned} P(X = x, Y = y) &= P(X + Y = x + y, X = x) \\ &= P(X + Y = x + y)P(X = x|X + Y = x + y) \\ &= \frac{e^{-\lambda}\lambda^{x+y}}{(x+y)!}P(X = x|X + Y = x + y) \end{aligned} \tag{2}$$

since $X + Y$ is the total number of accidents in the time interval and so has Poisson(λ) distribution.

Solution - Serious Accidents Ctd

Further given $X+Y = x+y$, the assumptions say that the serious accidents are the "successes" in $x + y$ Bernoulli trials with success probability p so

$$P(X = x|X + Y = x + y) = \binom{x+y}{x} p^x(1-p)^y$$

Putting this together with (2) gives the joint probability mass function as:

$$\begin{aligned} P(X = x, Y = y) &= \frac{e^{-\lambda}\lambda^{x+y}}{(x+y)!} \times \binom{x+y}{x} p^x(1-p)^y \\ &= \frac{e^{-p\lambda}e^{-(1-p)\lambda}\lambda^x\lambda^y}{(x+y)!} \times \frac{(x+y)!}{x!y!}p^x(1-p)^y \\ &= \frac{e^{-p\lambda}(p\lambda)^x}{x!} \times \frac{e^{-(1-p)\lambda}((1-p)\lambda)^y}{y!}. \end{aligned} \tag{3}$$

Solution - Serious Accidents Ctd 2

Remarkably the joint probability mass function is a product of Poisson($p\lambda$) and Poisson($(1-p)\lambda$) probabilities so the marginal probability mass function for X is given by

$$\begin{aligned}
P(X = x) &= \sum_{y=0}^{\infty} P(X = x, Y = y) \\
&= \sum_{y=0}^{\infty} \frac{e^{-p\lambda}(p\lambda)^x}{x!} \times \frac{e^{-(1-p)\lambda}((1-p)\lambda)^y}{y!} \\
&= \frac{e^{-p\lambda}(p\lambda)^x}{x!} \sum_{y=0}^{\infty} \frac{e^{-(1-p)\lambda}((1-p)\lambda)^y}{y!} \\
&= \frac{e^{-p\lambda}(p\lambda)^x}{x!}. \qquad (4)
\end{aligned}$$

Solution - Serious Accidents Ctd 3

Last step in (4) follows from property (b) for the pmf of Y and so the pmf of X is Poisson($p\lambda$).

Similarly , summing over x, the pmf of Y is Poisson($(1-p)\lambda$).

And because (3) shows that the joint pmf is the product of the two marginal pmf's, X and Y are independent.

Comment 1: surprising (?) that any information about the non-serious accidents in a period does not change the probabilities for the serious accidents in the period.

Comment 2: Whenever the joint pmf or pdf is a product of two pmf's or pdf's, then these pmf's or pdf's are the marginals and the two random variables are independent.

1.4 Continuous Joint PDF - Maximum and Minimum Example

Example - Minimum, Maximum of independent U(0,1)

A computer generates two independent random variables U and V with uniform distribution on (0,1). Let X be the minimum of U and V and Y be the maximum. What is the joint pdf and the marginal pdf's. Are X and Y independent?

Solution - Minimum, Maximum of Independent U(0,1)

Suppose U, V are the independent U(0,1) random variables.

Joint pdf of U, V is f where

$$f(u, v) = \begin{cases} 1, & 0 < u, v < 1 \\ 0, & \text{otherwise} \end{cases}$$

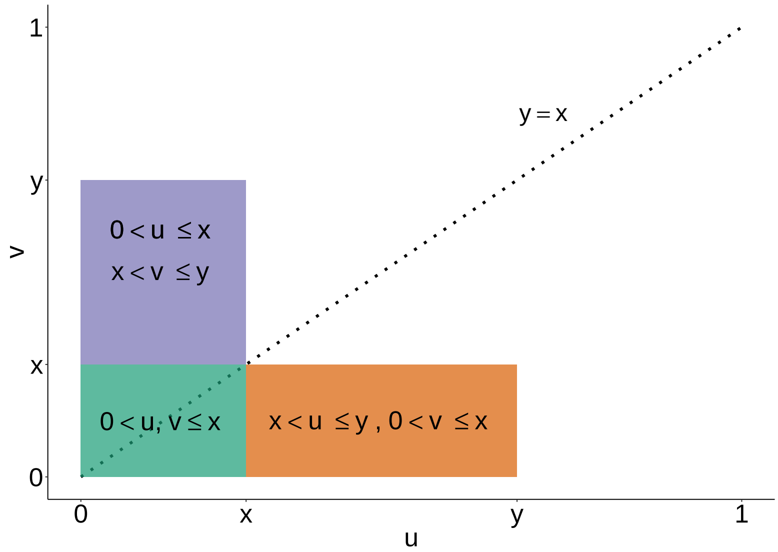

Figure 1: Shaded area is A so $P(X \leq x, Y \leq y) =$ area of A

Let X be the minimum of U and V, and Y be the the maximum of U and V.

By property (c) of joint pdf's,

$$P(X \leq x, Y \leq y) = \int \int_A f(u, v)\, dudv$$

where $A = \{(u, v) : \min(u, v) \leq x, \max(u, v) \leq y\}$.

Solution - Min, Max of Indep't U(0,1) Ctd

Since f is constant at 1 on the unit square, integrating f over A is the same as finding the area of A.

Figure 1 shows that A has three parts: green, brown and purple

$P(X \leq x, Y \leq y) = 0$ if $x > y$ or $x < 0$, so only need consider $0 < x < y < 1$. (Note $P(U = V) = 0$).

Green: has both u and $v \leq x$ so both min. and max. of u, v are $\leq x < y$.

Brown: has the minimum as $v \leq x$ and the maximum as $u \leq y$.

Purple: has the reverse: minimum as $u \leq x$ and the maximum as $v \leq y$.

Unshaded: part of unit square that has either the minimum $> x$ or the maximum $> y$ or both.

Solution - Min, Max of Indep't U(0,1) Ctd 2

So

$$\begin{aligned} P(X \leq x, Y \leq y) &= \text{area of green} + \text{area of purple} + \text{area of brown} \\ &= x^2 + (y-x)x + (y-x)x \\ &= 2yx - x^2 \end{aligned}$$

Differentiate wrt both x and y to get the joint pdf, $f_{(X,Y)}$, of (X,Y):

$$\begin{aligned} f_{(X,Y)}(x,y) &= \begin{cases} \frac{\delta}{\delta y}\frac{\delta}{\delta x}P(X \leq x, Y \leq y), & 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{\delta}{\delta y}(2y - 2x) = 2, & 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

Solution - Min, Max of Indep't U(0,1) Ctd 3

Integrate the joint density with respect to each variable to get the marginal densities of X and Y:

$$\begin{aligned} f_X(x) &= \begin{cases} \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\,dy, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} \int_x^1 2\,dy = 2(1-x), & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}. \end{aligned}$$

Solution - Min, Max of Indep't U(0,1) Ctd 4

And

$$\begin{aligned} f_Y(y) &= \begin{cases} \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\,dx, & 0 < y < 1 \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} \int_0^y 2\,dx = 2y, & 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Check: $P(Y \leq y) = P(\text{both}\, U, V \leq y) = y^2 = \int_0^y f_Y(w)\,dw$, as required. Check: $P(X > x) = P(\text{both}\, U, V > x) = (1-x)^2 = \int_x^1 f_X(w)\,dw$, as required.

Solution - Min, Max of Indep't U(0,1) Ctd 5

X and Y are *not* independent because the joint pdf is not the product of the marginals.

Can't be independent because the joint pdf is 0 for $x > y$ but the product of the marginals for $0 < y < x < 1$ is $4y(1-x)$.

Comment: Whenever the joint pdf is 0 on some region where both marginals are positive, there are pairs of values which are not possible in combination but the individual values are possible and the random variables can't be independent.

2 Covariance and Correlation - 4.2

2.1 Expectation of a function of two random variables

Expectation of a function of two random variables

Univariate: If X is a rv with pmf or pdf f and h is a function from reals to real, then

$$E(h(X)) = \begin{cases} \sum h(x)f(x), & \text{X discrete} \\ \int_{-\infty}^{\infty} h(x)f(x)\,dx, & \text{X continuous} \end{cases}.$$

The sum for X discrete is over the possible values of X - those that have prob. > 0 - called the range X. Bivariate: If X and Y have pmf or pdf f and h is a function from the plane to the plane, then

$$E(h(X,Y)) = \begin{cases} \sum h(x,y)f(x,y), & \text{X,Y discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)f(x,y)\,dxdy, & \text{X,Y continuous} \end{cases} \tag{5}$$

Similarly, for (X,Y) discrete, the sum is over the *pairs* (x,y) with prob. > 0 - the range of X,Y.

Expectation of a function of two random variables

Textbook calls the range in both cases the *support* of the random variable.

Illustration with linear combination of two random variables

If $h(x,y) = ax + by, a, b \in \mathbb{R}$ then, using equation (5),

$$\begin{aligned} E(aX+bY) &= E(h(X,Y)) \\ &= \begin{cases} \sum (ax+by)f(x,y), & \text{X,Y discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (ax+by)f(x,y)\,dxdy, & \text{X,Y continuous} \end{cases} \\ &= \begin{cases} a\sum xf(x,y) + b\sum yf(x,y) & \text{X,Y discrete} \\ a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,dxdy + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy, & \text{X,Y continuous} \end{cases} \\ &= \begin{cases} a\sum xf_X(x) + b\sum yf_Y(y), & \text{X,Y discrete} \\ a\int_{-\infty}^{\infty} xf_X(x)\,dxdy + a\int_{-\infty}^{\infty} yf_Y(y)\,dy, & \text{X,Y continuous} \end{cases} \\ &= aE(X) + bE(Y). \end{aligned}$$

This was demonstrated for discrete random variables using the sample space in Module 2 Section 2.6.

Variance of Sum

From Module 2 Section 4.1 , equation (26) showed that the variance of a sum of *any* random variables, X and Y can be written as

$$Var(X+Y) = Var(X) + Var(Y) + 2\left(E(XY) - E(X)E(Y)\right)$$

Covariance is defined as the expression in brackets:

$$\begin{aligned} Cov(X,Y) &= E(XY) - E(X)E(Y) \\ &= E\left((X-\mu_X)(Y-\mu_Y)\right) \end{aligned} \tag{6}$$

following the argument in the next slide.

Proof of Covariance Identity

Expand brackets to get:

$$\begin{aligned} E\left((X-\mu_X)(Y-\mu_Y)\right) &= E\left(XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y\right)\\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y \\ &= E(XY) - \mu_X\mu_Y. \end{aligned}$$

where the second step follows from linearity and the third from the definitions of μ_X, μ_Y.

Covariance and Variance of Sum

So in general

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

If X and Y are *independent*, then $Cov(X,Y) = 0$.

Correlation

Correlation of two random variables X and Y is the covariance of the standardised random variables.

So if μ_X, μ_Y the means and σ_X, σ_Y are the standard deviations

$$\rho = Corr(X,Y) = E(Z_X Z_Y), \tag{7}$$

where

$$Z_X = \frac{X-\mu_X}{\sigma_X}, \quad Z_Y = \frac{Y-\mu_Y}{\sigma_Y}. \tag{8}$$

Remarkable fact: $-1 \le \rho \le 1$.

If $Y = aX + b$ for some numbers a, b, then the correlation is -1 or 1.

Correlation - Intuition

If large values of X and Y occur together, and the same for small values, the product $Z_X Z_Y$ will be large.

If there is no relationship between the outcomes when large or small values occur, then pluses and minuses will cancel each other out giving an expectation close to 0.

Vice-versa: Large values of X occurring with small values of Y, and vice-versa will make the product $Z_X Z_Y$ negative

For ρ is close to ± 1 the joint probability mass function or probability lies close to a line - called the *regression* line and connected to lines of best fit - see pp. 144 and 145 which will be discussed in Module 6 (Ch 6.5)

Correlation and Independence

If X and Y are *independent*, then $Corr(X, Y) = 0$.

But the reverse is NOT always true - there are pairs of random variables which are not independent but have correlation 0 (see example on Lab/Workshop 6).

Example - Trinomial - pp. 139, 145

Suppose there are three possible outcomes for sales for a month, namely: B, meaning "below budget", E, meaning "meets budget", A, meaning "above budget". Sales in April, May and June are independent of one another and have probability of 0.5 of E and 0.25 each of B and A. Find the mean and variance of the number of months which meet or exceed budget. What is the correlation of the number of months which meet budget with the number of months which exceed budget?

Solution - Trinomial

Let N_B, N_E, N_A be the random variables which give the number of months for which sales are below, meet or above budget (respectively).

Then $N_B + N_E + N_A = 3$ since on each occasion exactly one of the three possibilities occurs.

Assumptions say that we have 3 Bernoulli trials in which success is above or meeting budget and failure is below budget.

So $N_E + N_A \sim Bin(3, 0.75)$ since the probability of meeting or being above budget is $0.5 + 0.25$.

Hence the mean of the number of months meeting or above budget is $E(N_E + N_A) = 3 \times 0.75 = 2.25$ and the variance is $Var(N_E + N_A) = 3 \times 0.75 \times 0.25 = 0.5625$.

Solution - Trinomial Ctd

For $x, y \in 0, 1, 2, 3$

$$P(N_E = x, N_A = y) = \frac{6}{x!y!(3-x-y)!} 0.5^x \times 0.25^y \times 0.25^{3-x-y}$$

since

$0.5^x \times 0.25^y \times 0.25^{3-x-y}$ is the probability of getting x months meeting budget and y months above budget and $3-x-y$ below budget in April, May and June in that order and

The number of ways to choose the months in order to get x meeting budget, y above budget and therefore $3-x-y$ below budget is 6 and there are $x!y!(3-x-y)!$ ways to order the months to get the same result (convince yourself of this by thinking about specific cases, for example $x = 3$, $x = 2, y = 1$ etc).

Solution - Trinomial Ctd 2

The pmf can be recorded in the following table whose entries are $f(x,y) = P(N_E = x, N_A = y)$

x \ y	0	1	2	3	$P(N_E = x)$
0	$^1/_{64}$	$^3/_{64}$	$^3/_{64}$	$^1/_{64}$	$^1/_8$
1	$^6/_{64}$	$^{12}/_{64}$	$^6/_{64}$	0	$^3/_8$
2	$^{12}/_{64}$	$^{12}/_{64}$	0	0	$^3/_8$
3	$^8/_{64}$	0	0	0	$^1/_8$
$P(N_A = y)$	$^{27}/_{64}$	$^{27}/_{64}$	$^9/_{64}$	$^1/_{64}$	1

Check: the row sums of the joint pmf give the marginal pmf of $N_E \sim Bin(3, 0.5)$ and the column sums of the joint pmf give the marginal pmf of $N_A \sim Bin(3, 0.25)$

Solution - Trinomial Ctd 3

So

$$E(N_E N_A) = 1 \times 1 \times \frac{12}{64} + 1 \times 2 \times \frac{6}{64} + 2 \times 1 \times \frac{12}{64} = \frac{48}{64} = \frac{3}{4}$$

Cov'ce

$$Cov(N_E, N_A) = E(N_E N_A) - E(N_E)E(N_A) = \frac{3}{4} - \frac{3}{2} \times \frac{3}{4} = -\frac{3}{8}$$

So N_A and N_E are not independent - why is it intuitive that the correlation is negative?

$$Corr(N_E, N_A) = \frac{Cov(N_E, N_A)}{\sqrt{Var(N_E)Var(N_A)}} = -\frac{0.375}{\sqrt{0.75 \times 0.5625}} = -0.5774$$

Solution - Trinomial Ctd 4

Comment -0.6 is quite a high negative correlation - reflecting the dependence between the number of months meeting and above budget.

3 Conditional Distributions - 4.3

Definition

If X and Y are random variables with joint pmf/pdf f and marginal pmf/pdf's f_X and f_Y , then the *conditional pmf/pdf of X given $Y = y$, $g(x|y)$,* is defined (provided $f_Y(y) > 0$) as:

$$g(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

Note: the conditional pmf/pdf is defined only for values in the range of Y BUT note that if Y is continuous $P(Y = y) = 0$.

Discrete case:

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

Example - Cherry Trees

31 Canadian cherry trees were classified in height and volume according to the following table:

Height \ Volume	< Aver'e	Aver'e	> Aver'e	Large	Row Sum
< Average	4	2	0	0	6
Average	4	10	4	0	18
> Average	2	2	2	1	7
Column Sum	10	14	6	1	31

A tree is selected at random from the 31. Find the joint pmf and marginal pmf's of the randomly selected tree's height and volume. Find the conditional pmf for the volume of the tree given that it is below or above average, or average in volume? Comment.

Solution - Cherry Trees

Let X be the height category and Y be the volume category of the randomly selected tree.

Joint pmf is obtained by dividing the entries in the table by 31 to get the joint pmf $P(X = x, Y = y)$ where x and y range over the categories of height and volume, since there are 31 equally likely outcomes for the random selection.

Marginal pmfs are obtained by dividing the row sums and the column sums by 31 for the same reason.

Next slide gives the resulting table.

Solution - Cheery Trees Ctd

x=h't \ y=vol.	< Aver'e	Aver'e	> Aver'e	Large	$P(X = x)$
< Average	$^4/_{31}$	$^2/_{31}$	0	0	$^6/_{31}$
Average	$^4/_{31}$	$^{10}/_{31}$	$^4/_{31}$	0	$^{18}/_{31}$
> Average	$^2/_{31}$	$^2/_{31}$	$^2/_{31}$	$^1/_{31}$	$^7/_{31}$
$P(Y = y)$	$^{10}/_{31}$	$^{14}/_{31}$	$^6/_{31}$	$^1/_{31}$	1

Solution - Cherry Trees Ctd 2

Conditional pmf's for Y given $X = x$ are the ratio of the joint pmf entries in the table to $P(X = x)$ which is the row sum. They are also the ratio of the number of trees in the height/volume combination to the total number of trees with that height, since the 31 cancels in numerator and denominator. The table shows the answers:

x=h't \ y=vol.	< Aver'e	Aver'e	> Aver'e	Large	$P(X=x)$
< Average	$^2/_3$	$^1/_3$	0	0	$^6/_{31}$
Average	$^2/_9$	$^5/_9$	$^2/_9$	0	$^{18}/_{31}$
> Average	$^2/_7$	$^2/_7$	$^2/_7$	$^1/_7$	$^7/_{31}$
$P(Y=y)$	$^{10}/_{31}$	$^{14}/_{31}$	$^6/_{31}$	$^1/_{31}$	1

Solution - Cherry Trees Ctd 3

Comment As the tree gets taller, so the volume increases.

Less than average height trees have average or below average volume.

Average height trees mostly have average volume .

Above average height trees are spread across the volume categories.

So: The random variables X and Y are not independent since the conditional distributions are not the same - note the 0's in the table prevent this.

4 Bivariate Normal Distribution - 4.4

Combination of indep't Normal rv's

If X and Y are independent random variables each with normal distribution, then, for any numbers a, b, c, the random variable

$$aX + bY + c \sim N(a\mu_X + b\mu_Y + c, (a\sigma_X)^2 + (b\sigma_Y)^2) \tag{9}$$

Combination of indep't Normal rv's - proof

Workshop/Lab 6 question 1(b) did one case of this with $a = 5, b = -2, c = 6$. In general: For any t, using independence of X and Y and linearity of expectation:

$$\begin{aligned} E(\exp(t(aX + bY + c)) &= \exp(tc)E(\exp(taX))E(\exp(tbY)) \\ &= \exp(tc)\exp\left[ta\mu_X + \frac{(ta\sigma_X)^2}{2}\right] \times \\ &\quad \exp\left[tb\mu_Y + \frac{(tb\sigma_Y)^2}{2}\right]. \\ &= \exp\left[t(c + a\mu_X + b\mu_Y) + \frac{t^2((a\sigma_X)^2 + (b\sigma_Y)^2)}{2}\right] \end{aligned}$$

as required.

Standard Bivariate Normal Definition

Suppose Z_1, Z_2 are independent N(0,1) ie have standard normal distributions.

For take ρ to satisfy $-1 \le \rho \le 1$ and put

$$X = Z_1, \quad Y = \rho Z_1 + \sqrt{1-\rho^2}Z_2. \tag{10}$$

Note: $X \sim N(0,1)$ and $Y \sim N(\rho \times 0 + \sqrt{1-\rho^2} \times 0, \rho^2 + (1-\rho^2)) = N(0,1)$.

Further: $Corr(X,Y) = E(XY) = E(\rho Z_1^2 + \sqrt{1-\rho^2} Z_1 Z_2)$.

But the independence of Z_1, Z_2 gives $E(Z_1 Z_2) = E(Z_1)E(Z_2) = 0$, so $Corr(X,Y) = \rho E(Z_1^2) = \rho$.

The pair (X,Y) is said to have a standard bivariate normal distribution with correlation ρ.

General Bivariate Normal Definition

Suppose we want means μ_X, μ_Y and standard deviations σ_X, σ_Y.

With the same restriction that $-1 \le \rho \le 1$, put

$$X = \mu_X + \sigma_X Z_1, \quad Y = \mu_Y + \rho\sigma_Y Z_1 + \sqrt{1-\rho^2}\sigma_Y Z_2. \tag{11}$$

Then $E(X) = \mu_X + \sigma_X \times 0 = \mu_X$, $E(Y) = \mu_Y + \rho\sigma_Y \times 0 + \sqrt{1-\rho^2}\sigma_Y \times 0 = \mu_Y$,

With $Var(X) = \sigma_X^2 \times 1 = \sigma_X^2$ and $Var(Y) = \rho^2\sigma_Y^2 \times 1 + (1-\rho^2)\sigma_Y^2 \times 1 = \sigma_Y^2$.

Define: the functions z_X, z_Y from reals to reals by

$$z_X(x) = \frac{x-\mu_X}{\sigma_X}, \quad z_Y(y) = \frac{y-\mu_Y}{\sigma_Y}, \tag{12}$$

So $z_X(X) = Z_1, z_Y(Y) = \rho Z_1 + \sqrt{1-\rho^2} Z_2$

Giving $Corr(X,Y) = E\left(z_X(X) z_Y(Y)\right) = \rho$.

Bivariate Normal Properties

Marginal distributions are normal from (9):

$$X \sim N(\mu_X, \sigma_X^2), Y \sim N(\mu_Y, \sigma_Y^2). \tag{13}$$

Given $Z_1 = z, Y = \mu_Y + \sigma_Y \rho z + \sigma_Y \sqrt{1-\rho^2} Z_2$ so equation (9) gives $Y|Z_1 = z \sim N(\mu_X + \sigma_X \rho z, (1-\rho^2)\sigma_Y^2)$.

Conditional pdf of Y given $X = x$ is the same as the conditional pdf of Y given $z_X(X) = z_X(x)$ (see (12)) and so

$$Y|X = x \sim N(\mu_Y + \rho\sigma_Y z_X(x), (1-\rho^2)\sigma_Y^2). \tag{14}$$

Regression to the Mean

$\rho = \pm 1$: the conditional pdf is that of $N(\mu_Y + \sigma_Y z_X(x), 0)$. Here there is *no* variation around the predicted value, $\mu_Y + \sigma_Y z_X(x)$.

$\rho = 0$: the conditional pdf is that of $N(\mu_Y, \sigma_Y^2)$ whatever is the value of x. So X gives *no* information on Y and the two random variables are thus *independent*.

Note: If (X,Y) is bivariate normal, correlation 0 is the same as independence. See Lab/Workshop 6 Question 26 for normal distributed rv's with correlation 0 which are *not* independent.

In between: there is shrinkage of the distribution variance by the square of the correlation coefficient and the mean value of the conditional distribution is brought back towards μ_Y from the value when $\rho = \pm 1$ - this is called regression to the mean.

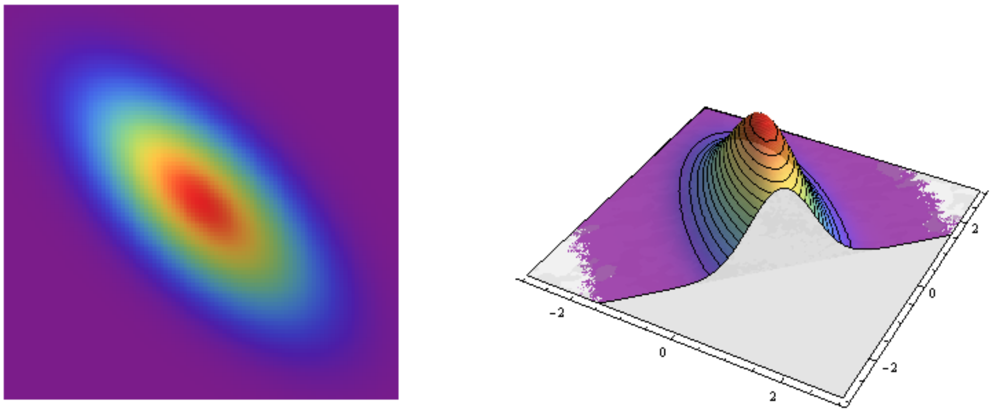

Figure 2: Contour Diagram of Bivariate Normal with Section of PDF Shown

Joint pdf of bivariate normal

By definition of the conditional pdf, the conditional pdf by the marginal pdf gives the joint pdf.

The exponent in the conditional pdf of $Y|X = x$ evaluated at y, using the standardising notation in (12), is

$$-\frac{(y - (\mu_Y + \rho\sigma_Y z_X(x))^2}{2\sigma_Y^2(1-\rho^2)} = -\frac{(z_Y(y) - \rho z_X(x))^2}{2(1-\rho^2)}.$$

So, after putting the sum of the marginal and conditional exponents on a common denominator $1 - \rho^2$, the joint density is:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times \exp\left[-\frac{z_X^2(x) - 2\rho z_X(x) z_Y(y) + z_Y^2(y)}{2(1-\rho^2)}\right]. \tag{15}$$

Example - Bivariate Normal

The scores which sit behind the tertiary entrance rank, ATAR, and the Year 10 performance test, NAPLAN, for Victorian students approximately have a bivariate normal distribution with correlation 0.7. If a student has a NAPLAN score 2 standard deviations above the mean, what is the probability that the ATAR score will be at least 2 standard deviations above the mean?

Solution - Bivariate Normal

Let Z_A, Z_N be the scores behind ATAR and NAPLAN for a randomly chosen Victorian student. Assumptions say that Z_A, Z_N have an approximate standard bivariate normal distribution with correlation 0.7. So the conditional distribution of Z_A given $Z_N = 2$ is $N(0 + 0.7 \times 2, (1 - 0.49)) = N(1.4, 0.51)$. Hence the required probability is the probability that this normal distribution is above 2 which is about 0.2 from the R command on the next slide:

Solution - Bivariate Normal

```
1 - pnorm(2, mean = 1.4, sd = sqrt(0.51))

## [1] 0.2004071
```

Only about 2.5% of students have values about 2 in a standard normal distribution so the NAPLAN score was in this 2.5%. However, because of regression to the mean and shrinkage of variance in the bivariate normal, only 20% of these students will go on to be in the top 2.5% of ATAR scores.