Univariate and Bivariate Distributions

1. Introduction to Univariate Distributions

For a random experiment, with a sample space, Ω, a function *X*, which assigns to each element of *C*, a real number *X*(*C*)=*x*, is called a random variable. We distinguish between the random variable, *X*, and the possible outcomes, $x \in \Omega$.

Example 1:

Consider rolling a dice then $\Omega = \{1,2,3,4,5,6\}$, define a random variable, which considers only odd or even numbers and $X(C) = \begin{cases} 1 & \text{if even number} \\ 0 & \text{if odd number} \end{cases}$, then

x	0	1
$P(X=x)$	½	½

2. Expectations and variances

X takes on a finite number of outcomes $x = x_1, x_2, \ldots, x_n$ and each has an associated probability:

X	x_1	x_2	…	x_n
$P(X=x)$	p_1	p_2	…	p_n

$$E(X) = \sum_x p_X(x)x = \mu_X$$

$$V(X) = E\left[(X-\mu_X)^2\right] = \sum_x p_X(x)(x-\mu_X)^2$$

$$= \sum_x p_X(x)x^2 - 2\mu_X \underbrace{\sum p_X(x)x}_{\mu_X} + \mu_X^2 \underbrace{\sum_x p_X(x)}_{1} = \sum_x p_X(x)x^2 - \mu_X^2 = E(X^2) - \mu_X^2$$

$$V(X) = E\left[(X-\mu_X)^2\right] = E(X^2) - E(X)^2$$

in general, $E\left[g(X)\right] = \sum_x p_X(x)g(x)$

for example,

$$E\left[X^2\right] = \sum_x p_X(x)x^2$$

Rules on expectations and variance for univariate distributions are in Appendix 1.

Example

x	1	2	3
Pr(.)	0.4	0.3	0.3

$$E(X) = 0.4(1) + 0.3(2) + 0.3(3) = 1.9$$

$$E(X^2) = 0.4(1^2) + 0.3(2^2) + 0.3(3^2) = 4.3,\ V(X) = E(X^2) - \left[E(X)\right]^2 = 4.3 - 1.9^2 = 0.69$$

3. Introduction to Discrete Bivariate Distributions

Let $p_{X_1,X_2}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$ be the joint probability density (mass) function for the discrete random variables X_1 and X_2. The joint probability density function defines a probability for all values of (x_1, x_2). Suppose that the sample space of X_1 is $\Omega_1 = \{A_1, A_2, A_3\}$ and the sample space of X_2 is $\Omega_2 = \{B_1, B_2\}$. Then we can define all possible outcomes and the joint probability density function using the Table 1 below:

Table 1: Joint Probability Table

	B_1	B_2	Total
A_1	$P(A_1 \cap B_1)$	$P(A_1 \cap B_2)$	$P(A_1)$
A_2	$P(A_2 \cap B_1)$	$P(A_2 \cap B_2)$	$P(A_2)$
A_3	$P(A_i \cap B_1)$	$P(A_i \cap B_2)$	$P(A_3)$
Total	$P(B_1)$	$P(B_2)$	1

The table defining a probability for each pair of events (joint probability), such that $P(A_i, B_j) \geq 0$ and $\sum_{x_1}\sum_{x_2} p_{X_1,X_2}(x_1, x_2) = 1$ then it is a valid probability density function.

We can also define a marginal probability, $\Pr(X_1 = A_2,$ X_2 takes any possible value) as

$$P(X_1 = A_2, B_1 \leq X_2 \leq B_2) = \sum_{x_2} p_{X_1,X_2}(A_2, x_2)$$

$$P(X_1 = A_2, B_1 \leq X_2 \leq B_2) = P(A_2 \cap B_1) + P(A_2 \cap B_2)$$

So $\sum_{x_2} p(x_1, x_2)$ removes the variable X_2 out of the probability formula, leaving the marginal probability, denoted as $p_1(x_1)$, and this is a probability density function of X_1, x_2 having been summed out, and is known as the MARGINAL PROBABILITY DENSITY (MASS) FUNCTION of X_1.

From these marginal distributions we can calculate the moments of the random variables X_1 and X_2, that is, $E(X_1)$, $E(X_2)$, $V(X_1)$, $V(X_2)$ and $\text{cov}(X_1, X_2)$ as we did before. For the discrete random variable X_1, with marginal probability density function $p_1(x_1)$:

$$E(X_1) = \sum_{x_1} x_1 p_1(x_1)$$

$$V(X_1) = \sum_{x_1} x_1^2 p_1(x_1) - E(X_1)^2$$

and

$$\text{cov}(X_1, X_2) = E(X_1 - E(X_1))(X_2 - E(X_2)) = E(X_1 X_2) - E(X_1)E(X_2)$$

$$\text{cov}(X_1, X_2) = \sum_{x_2}\sum_{x_1} x_1 x_2 p(x_1, x_2) - E(X_1)E(X_2)$$

Rules on expectations and variance for bivariate and higher order distributions are in Appendix 2.

3.2 Conditional Distributions

The conditional probability density function for random variables, X_1 and X_2 with a joint probability density (mass) function $p(x_1,x_2)$ and marginals $p_1(x_1)$ and $p_2(x_2)$, is written as:

$$p(x_1 \mid x_2) = \frac{p(x_1, x_2)}{p_2(x_2)}.$$

This is a valid p.d.f. as

$$\sum_{x_1} p(x_1 \mid x_2) = \sum_{x_1} \frac{p(x_1, x_2)}{p_2(x_2)} = \frac{1}{p_2(x_2)} \sum_{x_1} p(x_1, x_2) = \frac{1}{p_2(x_2)} \sum_{x_1} p(x_1, x_2) = \frac{p_2(x_2)}{p_2(x_2)} = 1$$

and $p(x_1 \mid x_2) = \dfrac{p(x_1, x_2)}{p_2(x_2)} > 0$

Example

		X_1			
		1	2	3	
	1	0.1	0.1	0.0	0.2
X_2	2	0.2	0.0	0.2	0.4
	3	0.1	0.2	0.1	0.4
		0.4	0.3	0.3	1.000

$E(X_1) = 1.9; V(X_1) = 0.09$

$E(X_2) = 0.2(1) + 0.4(2) + 0.4(3) = 2.2;$

$E(X_2^2) = 0.2(1^2) + 0.4(2^2) + 0.4(3^2) = 5.4 \Rightarrow V(X_2) = 0.56$

$E(X_1.X_2) = 0.1(1)(1) + 0.1(1)(2) + 0.0(1)(3) + 0.2(2)(1) + 0.0(2)(2) + 0.2(2)(3) +$
$0.1(3)(1)+0.2(3)(2)+0.1(3)(3)=4.0$

$\text{cov}(X_1, X_3) = E(X_1 X_3) = E(X_1).E(X_3) = 4.0 - 1.9(2.2) = -0.18$

Below we report the probability distribution of X_1 conditional on X_2=3:

$x_1 \mid X_2 = 3$	1	2	3
Pr(.)	0.25	0.5	0.25

$E(X_1 \mid X_2 = 3) = 0.25(1) + 0.5(2) + 0.25(3) = 1.5;$

$E(X_1^2 \mid X_2 = 3) = 0.25(1^2) + 0.5(2^2) + 0.25(3^3) = 4.5;\ \ V(X) = 4.5 - 1.5^2 = 2.25$.

Appendix 1: Rules on expectations variances

Define $E(w)=\sum_{i=1}^{k} p_i w_i$ as the expected value of the random variable w and

$V(w)=E[w-E(w)]^2=E(w^2)-E(w)^2=\sum_{i=1}^{k} p_i(w_i-E(w))^2$.

Then:

1. $E(a+x)=\sum_{i=1}^{k} p_i(a+x_i)=a+\sum_{i=1}^{k} p_i x_i=a+E(x)$.

2. $E(ax)=\sum_{i=1}^{k} p_i(ax_i)=a\sum_{i=1}^{k} p_i x_i=aE(x)$.

3. $V(a+x)=\sum_{i=1}^{k} p_i\left[(a+x_i)-E(a+x)\right]^2=\sum_{i=1}^{k} p_i\left[(a+x_i)-a-E(x)\right]^2$

   $V(a+x)=\sum_{i=1}^{k} p_i\left[x_i-E(x)\right]^2=V(x)$.

4. $V(ax)=\sum_{i=1}^{k} p_i\left[ax_i-E(ax)\right]^2=\sum_{i=1}^{k} p_i\left[ax_i-aE(x)\right]^2$

   $V(ax)=a^2\sum_{i=1}^{k} p_i\left[x_i-E(x)\right]^2=a^2V(x)$.

5. $E(x+y+z)=E(x)+E(y)+E(z)$.

6. $\text{cov}(a+x,y)=\text{cov}(x,y)$.

7. $\text{cov}(ax,y)=a\,\text{cov}(x,y)$.

8. $V(x+y)=V(x)+V(y)+2\,\text{cov}(x,y)$.

9. $V(x-y)=V(x)+V(y)-2\,\text{cov}(x,y)$.

10 $V(x+y+z)=V(x)+V(y)+V(z)+2\,\text{cov}(x,y)+2\,\text{cov}(x,z)+2\,\text{cov}(y,z)$

11 $V(x-y-z)=V(x)+V(y)+V(z)-2\,\text{cov}(x,y)-2\,\text{cov}(x,y)+2\,\text{cov}(y,z)$

Appendix 2: More rules on expectations variances

Define the probability density function for the random variables *X*, *Y* as $p(X,Y)$, such that, $\sum_x \sum_y p(x,y) = 1$. Define the marginal density of *X* as *p*(*x*) and the marginal probability density for *Y* as *p*(*y*), such that $p(X) = \sum_y p(x,y)$ and $p(Y) = \sum_x p(x,y)$.

Then define $E(X) = \sum_y xp(y)$, $E(Y) = \sum_y yp(y)$, $V(X) = \sum_x (x - E(X))^2 p(x)$, $V(Y) = \sum_y (y - E(Y))^2 p(y)$ and $\text{cov}(X,Y) = \sum_x \sum_y (x - E(X))(y - E(Y)) p(x,y)$.

Then we can show that

1. $E(X+Y) = \sum_x \sum_y (x+y) p(x,y) = \sum_x \sum_y xp(x,y) + \sum_x \sum_y yp(x,y)$

$$\sum_x x \underbrace{\sum_y p(x,y)}_{p(x)} + \sum_y y \underbrace{\sum_x p(x,y)}_{p(y)} = \sum_x xp(x) + \sum_y yp(y) = E(X) + E(Y)$$

2. $\text{cov}(a+X,Y) = \sum_x \sum_y ((a+x) - E(a+X))(y - E(Y)) p(x,y)$

$$= \sum_x \sum_y (a + x - a - E(X))(y - E(Y)) p(x,y) = \text{cov}(X,Y).$$

3. $\text{cov}(aX,Y) = \sum_x \sum_y (ax - E(aX))(y - E(Y)) p(x,y)$

$$\text{cov}(aX,Y) = \sum_x \sum_y (ax - aE(X))(y - E(Y)) p(x,y)$$

$$\text{cov}(aX,Y) = a \sum_x \sum_y (x - E(X))(y - E(Y)) p(x,y) = a\,\text{cov}(X,Y)$$

4. $V(X+Y) = \sum_x \sum_y \left[(x+y) - E(X+Y)\right]^2 p(x,y) = \sum_x \sum_y \left[(x - E(X)) + (y - E(Y))\right]^2 p(x,y)$

$$\sum_x \sum_y \left[(x - E(X)) + (y - E(Y))\right]^2 p(x,y)$$

$$\sum_x \sum_y (x - E(X))^2 p(x,y) + \sum_x \sum_y (y - E(Y))^2 p(x,y) + 2 \sum_x \sum_y (x - E(X)(y - E(Y)) p(x,y)$$

$$= \sum_x (x - E(X))^2 \underbrace{\sum_y p(x,y)}_{p(x)} + \sum_x (y - E(Y))^2 \underbrace{\sum_y p(x,y)}_{p(y)} + 2\,\text{cov}(X,Y)$$

$$= V(X) + V(Y) + 2\,\text{cov}(X,Y)$$

5. $E(X-Y) = E(X) - E(Y)$
6. $V(X-Y) = V(X) + V(Y) - 2\,\text{cov}(X,Y)$
7. $E(aX - bY) = aE(X) - bE(Y)$

8. $V(aX - bY) = a^2V(X) + b^2V(Y) - 2ab\operatorname{cov}(X,Y)$

9. $E(X+Y+Z) = E(X) + E(Y) + E(Z)$

10. $V(X+Y+Z) = V(X) + V(Y) + V(Z) + 2\operatorname{cov}(X,Y) + 2\operatorname{cov}(X,Z) + 2\operatorname{cov}(Y,Z)$

11. $E(X-Y-Z) = E(X) - E(Y) - E(Z)$

12. $V(X-Y-Z) = V(X) + V(Y) + V(Z) - 2\operatorname{cov}(X,Y) - 2\operatorname{cov}(X,Z) + 2\operatorname{cov}(Y,Z)$

Suppose $E(X_i) = \mu$, $V(X_i) = \sigma^2$ for all *i*.

Now define $\bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n} = \dfrac{(X_1 + X_2 + \ldots + X_n)}{n}$

$$E(\bar{X}) = E\left[\frac{(X_1 + X_2 + \ldots + X_n}{n}\right] = \frac{1}{n}E(X_1 + X_2 + \ldots + X_n)$$

$$= \frac{1}{n}\left[E(X_1) + \ldots + E(X_n)\right] = \frac{1}{n}\left[\mu + \mu + \ldots + \mu\right] = \mu$$

$$V(\bar{X}) = V\left[\frac{(X_1 + X_2 + \ldots + X_n)}{n}\right] = \frac{1}{n^2}[V(X_1) + V(X_2) + \ldots + V(X_n)$$
$$+2\operatorname{cov}(X_1, X_2) + 2\operatorname{cov}(X_1, X_3) + \ldots + 2\operatorname{cov}(X_1, X_n)$$
$$+2\operatorname{cov}(X_2, X_3) + \ldots + 2\operatorname{cov}(X_2, X_n)$$
$$+\ldots+$$
$$\ldots + 2\operatorname{cov}(X_{n-1}, X_n)]$$

assuming $X_1, \ldots, X_n$ are a random sample (with or without replacement, providing *n* is sufficiently large) then $\operatorname{cov}(X_{,}, X_j) = 0$ for $i \neq j$ and

$V(\bar{X}) = \dfrac{1}{n^2}[\sigma^2 + \sigma^2 + \ldots + \sigma^2] = \dfrac{\sigma^2}{n}$ (therefore the variance of the sample mean falls as the number of points in the sample increases).