Discrete Mathematics and Probability

Lecture 17
Transformations and Joint Probability

Ian Stark

School of Informatics
The University of Edinburgh

Thursday 14 November 2024

Week 10 Topics

Chapter 3: Continuous Random Variables and Probability Distributions

§3.7 Transformations of a Random Variable.

Chapter 4: Joint Probability Distributions and Their Applications

§4.1 Jointly Distributed Random Variables

§4.2 Expected Values, Covariance, and Correlation

§4.3 Properties of Linear Combinations (but not Moment Generating Functions)

The study guide and accompanying videos indicate the examinable course content. The corresponding sections in the book are to support this. Additional sections in the book are useful extension material but not required for this course.

Transforming a Random Variable

Let X be a continuous random variable. We have already seen situations where we work with a function of X rather than with X itself. One example is rescaling a random variable.

Rescaling

For any constants a and b the mean, variance and standard deviation of $(aX+b)$ can be calculated from the corresponding values for X.

$$E(aX+b) = aE(X)+b \qquad \mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(x) \qquad \mathrm{SD}(aX+b) = |a|\,\mathrm{SD}(X)$$

Here $(aX+b)$ is itself a continuous random variable, and we can in general consider random variables $Y=g(X)$ for arbitrary functions g.

Example Transformations

Let X be a continuous random variable with PDF $f_X(x)$ and CDF $F_X(x)$.

Suppose $Y = g(X)$ is a transformation giving another continuous random variable Y.

Example Transformations

- If X is a mass in kilograms, then $1000X$ is the same mass in grams.
- If R is the radius of a circle, then πR^2 is its area.
- If T is the time to travel a distance d, then (d/T) is the average speed for the journey.

We can ask: what are the PDF and CDF for these transformed variables?

Transformed CDF

Let X be a continuous random variable with PDF $f_X(x)$ and CDF $F_X(x)$.

Suppose $Y = g(X)$ is a transformation giving another continuous random variable Y with PDF $f_Y(y)$ and CDF $F_Y(y)$.

Suppose also that g is monotonically increasing: for all possible values $a < b$ of random variable X we have $g(a) < g(b)$.

Then there will be an inverse function h where $X = h(Y)$ and we can calculate as follows.

$$F_Y(y) \;=\; P(Y \leqslant y) \;=\; P(g(X) \leqslant y) \;=\; P(X \leqslant h(y)) \;=\; F_X(h(y))$$

If g is monotonically decreasing then it still has an inverse h but instead:

$$F_Y(y) \;=\; P(Y \leqslant y) \;=\; P(g(X) \leqslant y) \;=\; P(X \geqslant h(y)) \;=\; 1 - F_X(h(y))\,.$$

Example Transformations

Let X be a continuous random variable with PDF $f_X(x)$ and CDF $F_X(x)$.

Suppose $Y = g(X)$ is a transformation giving another continuous random variable Y, with PDF $f_Y(y)$ and CDF $F_Y(y)$, and where g is monotonically increasing or decreasing on the range of possible values for X.

Example Transformations

Grams and kilograms	$Y = 1000X$	$F_Y(y) = F_X(y/1000)$	
Area and radius	$A = \pi R^2$	$F_A(a) = F_R(\sqrt{a/\pi})$	$a \geqslant 0$
Speed and Time	$S = d/T$	$F_S(s) = 1 - F_T(d/s)$	

Example

1. Continuous random variable M is a program runtime in minutes, and H is the same runtime in hours. Give the function for calculating H from M.

   Give the inverse function, for calculating M from H.

2. F is a daily temperature on the Fahrenheit scale, and C is that temperature on the Celsius scale. Give the function for calculating C from F.

   Give the inverse function, for calculating F from C.

3. If $F_M(m)$ and $F_F(f)$ are cumulative distribution functions for M and F then give the corresponding CDFs $F_H(h)$ and $F_C(c)$ for H and C.

Transformed PDF

Let X be a continuous random variable with PDF $f_X(x)$.

Suppose $Y = g(X)$ is a transformation giving another continuous random variable Y, with PDF $f_Y(y)$.

Suppose that g is monotonic on the set of all possible values X.

Then there will be an inverse function $X = h(Y)$.

Suppose also that h has derivative $h'(y)$ for all the possible values of Y. Then we can directly calculate the PDF for Y.

$$f_Y(y) \;=\; f_X(h(y)) \cdot \left| h'(y) \right|$$

Example Transformations

Let X be a continuous random variable with PDF $f_X(x)$ and CDF $F_X(x)$.

Suppose $Y = g(X)$ is a transformation giving another continuous random variable Y.

Example Transformations

Grams and kilograms	$Y = 1000X$	$f_Y(y) = f_X(y/1000)/1000$	
Area and radius	$A = \pi R^2$	$f_A(a) = f_R(\sqrt{a/\pi})/(2\sqrt{\pi a})$	$a \geqslant 0$
Speed and Time	$S = d/T$	$f_S(s) = f_T(d/s) \cdot (d/s^2)$	

Example

4. Continuous random variable M with PDF $f_M(m)$ is a program runtime in minutes, and H is the same runtime in hours. Find the PDF $f_H(h)$ for H.

5. If continuous random variable F, measuring daily temperature on the Fahrenheit scale, has PDF $f_F(f)$, then calculate the PDF $f_C(c)$ for the random variable C giving the same temperature on the Celsius scale.

Week 10 Topics

Chapter 3: Continuous Random Variables and Probability Distributions

§3.7 Transformations of a Random Variable.

Chapter 4: Joint Probability Distributions and Their Applications

§4.1 Jointly Distributed Random Variables

§4.2 Expected Values, Covariance, and Correlation

§4.3 Properties of Linear Combinations (but not Moment Generating Functions)

The study guide and accompanying videos indicate the examinable course content. The corresponding sections in the book are to support this. Additional sections in the book are useful extension material but not required for this course.

Two Discrete Random Variables

Suppose that we have an experiment with sample space $\mathcal{S}$ and two discrete random variables X and Y both defined over that sample space. That is, for any outcome $s \in \mathcal{S}$ we have real numbers $X(s)$ and $Y(s)$. Then we can define the joint probability of the two random variables.

Definition

The *joint probability mass function* (JPMF) of X and Y is a function $p(x, y)$ defined for each possible pair (x, y) where X may take the value x and Y may take the value y.

$$p(x, y) \;=\; P(X = x \text{ and } Y = y)$$

For any set of pairs $A \subseteq \mathbb{R} \times \mathbb{R}$ the probability that (X, Y) lies in A is a sum over pairs.

$$P((X, Y) \in A) \;=\; \sum_{(x,y) \in A} p(x, y)$$

Marginal Probabilities

If we know the joint probability mass function $p(x, y)$ of X and Y then we can calculate the PMF of each variable individually.

Definition

The random variables X and Y have *marginal probability mass functions* $p_X(x)$ and $p_Y(y)$ given by summation.

$$p_X(x) = \sum_y p(x, y) \qquad p_Y(y) = \sum_x p(x, y)$$

Two Continuous Random Variables

Suppose that we have two continuous random variables X and Y both defined over the same sample space $\mathcal{S}$. That is, for any outcome $s \in \mathcal{S}$ we have real numbers $X(s)$ and $Y(s)$.

Definition

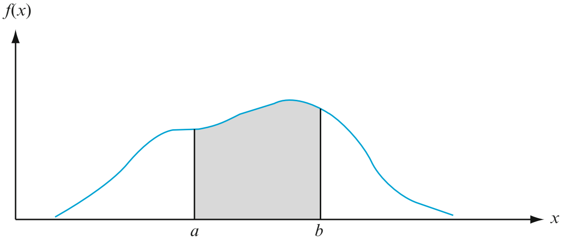
The *joint probability density function* (JPDF) of X and Y is a function $f(x, y)$ such that for any rectangle $A = \{ (x, y) \mid a \leqslant x \leqslant b, c \leqslant y \leqslant d \}$ we have the following.

$$P((X,Y) \in A) \; = \; P(a \leqslant X \leqslant b, c \leqslant Y \leqslant d) \; = \int_a^b \left(\int_c^d f(x,y)\,dy \right) dx \; = \int_c^d \left(\int_a^b f(x,y)\,dx \right) dy$$

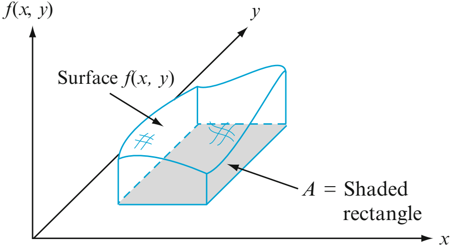
(Note: Although this result extends to an arbitrary region R using a "double integral" $\iint_R$, the construction here for rectangle A is the technically simpler *iterated integral*.)

Two Continuous Random Variables

If we treat the joint probability density function $f(x, y)$ as giving a height above the point (x, y) in a three-dimensional coordinate system then we get a surface in the same way that a single-variable PDF gives a curve.



Carlton & Devore Figure 3.2



Carlton & Devore Figure 4.1

For the curve on the left, probability is given by the area under the curve. For the surface on the right, probability is given by the volume under the surface.

For rectangle A on the right the volume under the surface is the probability $P((X, Y) \in A)$.

Marginal Probabilities

Again we can calculate the probability density functions for individual random variables from their joint PDF.

Definition

Continuous random variables X and Y have *marginal probability density functions* $f_X(x)$ and $f_Y(y)$ given by integration.

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \qquad\qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

Often the values of X or Y will be known to lie within a particular interval, with probability density zero outside. It's then possible to restrict the range of integration to just that interval.

Independent Random Variables

In many situations two random variables X and Y will be linked: knowing information about one will tell us about the other. For example, the transformations of random variables given earlier have this: if $Y = g(X)$ then knowing the value of X immediately tells us the value of Y.

Sometimes, though, variables are unconnected and the value of X says nothing about that of Y.

Definition

Two random variables are *independent* if for every pair of values x and y we have

$$p(x, y) = p_X(x) \cdot p_Y(y) \quad \text{for discrete random variables } X \text{ and } Y \text{, or, equivalently,}$$

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad \text{for continuous random variables } X \text{ and } Y.$$

Alternatively, if these equations fail for some (x, y) then X and Y are *dependent*.

This is similar to events A and B being independent if $P(A \cap B) = P(A) \cdot P(B)$.

More Than Two Random Variables

These ideas extend further, from two random variables being jointly distributed to multiple random variables over the same sample space.

Definition

If $X_1, X_2, \ldots, X_n$ are all discrete random variables then their *joint probability mass function* (JPMF) is

$$p(x_1, x_2, \ldots, x_n) \;=\; P(\, X_1 = x_1 \,\cap\, X_2 = x_2 \,\cap \cdots \cap\, X_n = x_n \,)\,.$$

If these are all continuous random variables then their *joint probability density function* (JPDF) is such that for n intervals $[a_1, b_1], [a_2, b_2], \ldots, [a_n, b_n]$ we have

$$P(a_1 \leqslant X_1 \leqslant b_1, \ldots, a_n \leqslant X_n \leqslant b_n) \;=\; \int_{a_1}^{b_1} \left(\ldots \left(\int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_n \right) \ldots \right) dx_1\,.$$

Optional Exercise: Not Examinable, Not Assessed

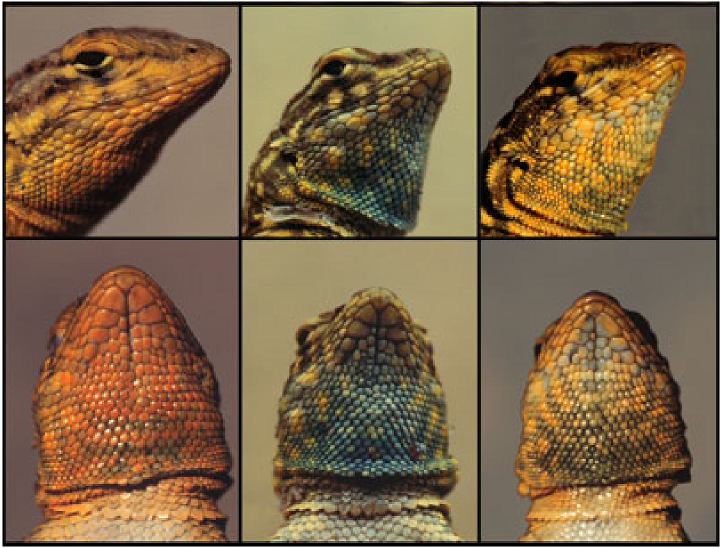
I have three six-sided dice that roll fairly but have unusual numbering.

Red	4	4	4	4	4	9
Green	0	5	5	5	5	5
Blue	2	2	2	7	7	7

Calculate the possible outcomes and their probabilities when rolling two dice, one red and one green. Which is more likely to get the highest score?

Do the same for a green die and a blue die rolled at the same time; and again for blue rolled against red.

A friend suggests playing a game where you get to choose any one of the three dice, they choose another, you both roll them and whoever scores highest wins. What should you do?



Picture Credit: Barry Sinervo, UCSC
https://bio.research.ucsc.edu/~barrylab/lizardland/male_lizards.overview.html

Summary

Topics

- Transforming a Random Variable
- Transformation of Cumulative Distribution Function
- Transformation of Probability Density Function
- Joint probability mass function (JPMF)
- Joint probability density function (JPDF)
- Independent random variables
- Marginal probability mass function (MPMF)
- Marginal probability density function (MPDF)
- More than two random variables

Reading

Chapter 3, §3.7; pp. 216–220. Chapter 4, §4.1; pp. 239–249.

Exercises

Chapter 3, Exercises 112–128; pp. 220–221. Chapter 4, Exercises 1–21; pp. 249–254.