STAT115: Introduction to Biostatistics

University of Otago
Ōtākou Whakaihu Waka

Lecture 9: More on Random Variables

Outline

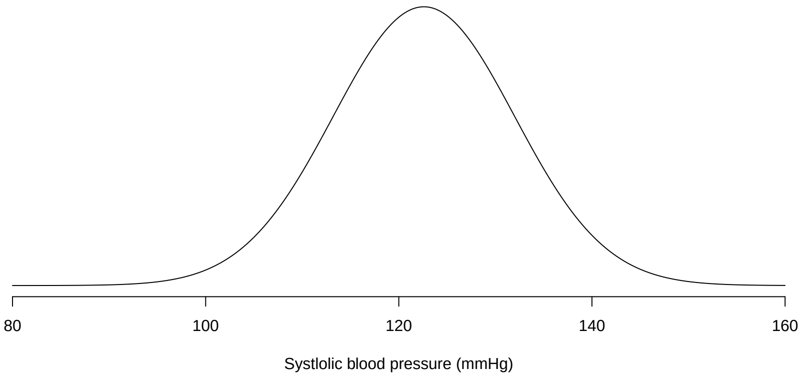
- We saw random variables in the last lecture
  - Probability distribution
  - Expectation
  - Variance
- Continue learning about random variables today
  - Can we have continuous random variables?
  - What happens when we combine random variables?

Discrete vs continuous

- The random variables we looked at in the last lecture were all discrete
  - Countable number of distinct values
- Discrete random variables are useful in a range of problems, e.g.
  - Number of eggs in a nest
  - Number of tasks completed in fixed time
  - Number of bugs in a piece of computer code
  - Number of voters who prefer National
- There are other situations where things aren't discrete, e.g.
  - The systolic blood pressure of a patient
  - The time taken in reflex test
  - The pH of seawater
- These can take continuous values

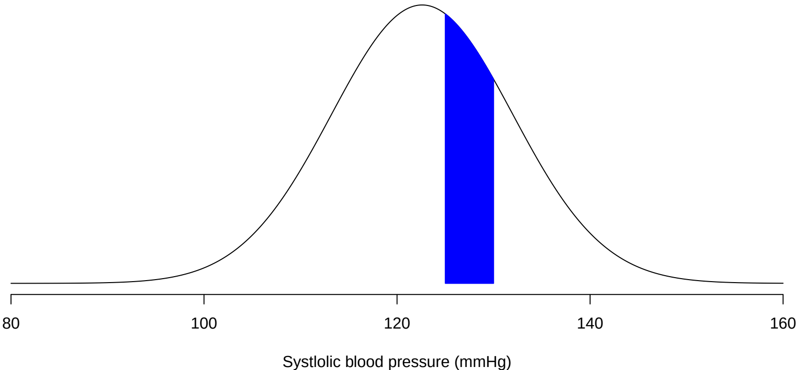
Continuous random variables

- An infinite (and uncountable) number of possible values
- Each value has a probability density
  - Best seen graphically (e.g. systolic blood pressure for females)



Probability density

- This curve is called a probability density function (pdf)
- Probability is given by the area under the curve (pdf)
  - The total area under the curve (pdf) is 1
- The probability of systlolic blood pressure 125 and 130 mmHg is given by:



Continuous vs discrete

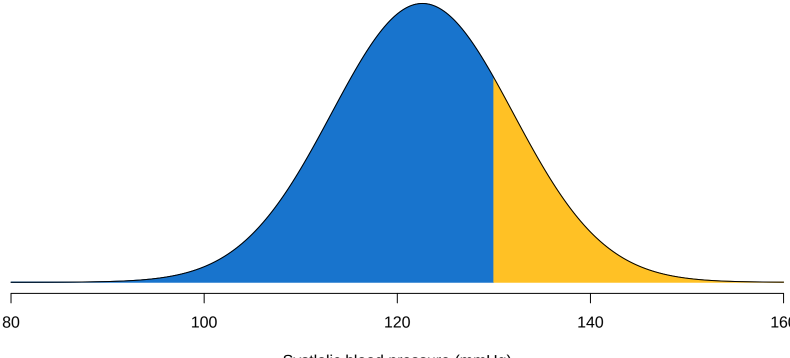
- Much of what we have already learned applies to continuous random variables
  - We can find expectation, variance, standard deviation
  - The calculations are more complex (sums are replaced by integrals)
  - Explore in more detail in more advanced courses (e.g. STAT 270)

Continuous

- Much of what we have already learned applies to continuous random variables
  - We can find expectation, variance, standard deviation
  - The calculations are more complex (sums are replaced by integrals)
  - Explore in more detail in more advanced courses (e.g. STAT 270)
- Look at examples on the next two slides

Complement

- Suppose we know the probability that systolic blood pressure is less than 130 mm (blue)
    - $\Pr(\text{systolic pressure} < 130) = 0.78$
- What is $\Pr(\text{systolic pressure} > 130)$? (gold)
    - It is a complement!



Combinations of random variables

- We may be interested in the combination of several random variables
  - Cholesterol
    - Random variables: (i) HDL cholesterol, (ii) LDL cholesterol, (iii) triglycerides
    - Combination: total cholesterol
  - Genetic linkage (crossover[1])
    - Random variables: number of crossovers in each chromosome
    - Combination: total number of crossovers
  - Cricket: runs scored
    - Random variables: number of singles, twos, threes, fours, sixes in an innings.
    - Combination: total score
  - Finance: portfolio value
    - Random variables: share prices for spark (SPK) and port of Tauranga (POT)
    - Combination: portfolio value (e.g. portfolio: 5 SPK, 10 POT)

[1] segments of DNA from one parent's chromosome swap with corresponding segments on the other parent's chromosome during meiosis

Combination of random variables

- Suppose we have random variables X and Y
  - To guide the development, we will think about
    - X: value of one SPK share in one months time
    - Y: value of one POT share in one months time
- We may be interested in a linear combination of X and Y
  - $aX + bY$
- What is the expected value of $aX + bY$?
- What is the variance of $aX + bY$

Expected value of combination

- If we owned shares: 5 SPK and 10 POT
  - Linear combination represents the value of our portfolio in one months time
  - $5X + 10Y$
    - Here, a is the number of SPK shares: 5
    - Here, b is the number of POT shares: 10
- How do we find the expected value of the linear combination?

$$E[aX + bY] = aE[X] + bE[Y]$$

- If $E[X] = 3$ and $E[Y] = 6.3$ then, the expected portfolio value is

$$\begin{aligned} E[5X + 10Y] &= 5E[X] + 10E[Y] \\ &= 5 \times 3 + 10 \times 6.3 \\ &= 78 \end{aligned}$$

Expected value of combination

- Ice cream is sold from 16 L containers in NZ
  - Expect that there is 16 L when opened
  - Can vary: let's say a standard deviation of 0.1 L (variance 0.01)
  - Let X be the amount of ice cream in a container: $E[X] = 16$, $Var(X) = 0.01$
- A new container of goldrush icecream is opened for the person ahead of us in line.
- They get a scoop of gold rush
  - Expect each scoop to get 0.1 L of ice cream
  - Standard deviation of 0.01 L (variance 0.0001).
  - Let Y be the amount in a scoop of ice cream: $E[Y] = 0.1$, $Var(Y) = 0.0001$
- The amount of goldrush icecream when we come to order is $X - Y$
  - What is $E[X - Y]$?

Variance of combination

- Can also be important to have a measure of variability for the combination of random variables
  - Total cholesterol
  - Number of crossovers
  - Runs in cricket innings
  - Value of portfolio
- If X and Y are independent, then

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

- If X and Y are not independent
  - The variance is more complicated (additional term needed)
  - Considered in higher level courses

Variance of combination

- What is $Var(X - Y)$ for ice cream example?
  - $a = 1$
  - $b = -1$

$$\begin{aligned} Var(X - Y) &= 1^2 Var(X) + (-1)^2 Var(Y) \\ &= Var(X) + Var(Y) \\ &= 0.01 + 0.0001 \\ &= 0.0101 \end{aligned}$$

- Portfolio: what is $Var(5X + 10Y)$?
  - Assume that share prices are independent (unlikely to be the case in reality)

Variance of combination

- We saw that $Var(X - Y) = Var(X) + Var(Y)$
  - We are subtracting Y from X. Why do the variances add?
- A server with low variability
  - Each scoop has is consistent in terms of the amount of ice cream
- A server with high variability
  - Each scoop can vary greatly (small or large or anywhere in between)
- If server is highly variable, will amount left in container be highly variable?

Variance of combination

- We saw that $Var(X - Y) = Var(X) + Var(Y)$
  - We are subtracting Y from X. Why do the variances add?
- A server with low variability
  - Each scoop has is consistent in terms of the amount of ice cream
- A server with high variability
  - Each scoop can vary greatly (small or large or anywhere in between)
- If server is highly variable, will amount left in container be highly variable?
- The variability in the amount of ice cream is the same if:
  - Add a scoop of ice cream to the container, or
  - Took a scoop of ice cream away

Abstract example

- Look at another example: somewhat abstract
  - Provide some useful results that we will use in coming weeks
- Let Y_1 and Y_2 be independent observations from a distribution
  - Mean μ
  - Standard deviation σ
- What is the mean and variance of $\frac{Y_1+Y_2}{2}$?
  - Sample mean of two values from a distribution

Abstract example: expected value

- The expected value of the sample mean is

$$E\left[\frac{Y_1 + Y_2}{2}\right] = \frac{1}{2}E[Y_1] + \frac{1}{2}E[Y_2]$$
$$= \frac{1}{2}\mu + \frac{1}{2}\mu$$
$$= \mu$$

- The variance of the sample mean is

$$Var\left(\frac{Y_1 + Y_2}{2}\right) = \frac{1}{4}Var(Y_1) + \frac{1}{4}Var(Y_2)$$
$$= \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2$$
$$= \frac{\sigma^2}{2}$$

Abstract example: extension

- This can be extended to when we have n independent observations: $Y_1, Y_2, \ldots, Y_n$
- The expected value of the sample mean is

$$\begin{aligned} E\left[\frac{Y_1 + Y_2 + \ldots + Y_n}{n}\right] &= \frac{1}{n}E[Y_1] + \frac{1}{n}E[Y_2] + \ldots + \frac{1}{n}E[Y_n] \\ &= \mu \end{aligned}$$

- The variance of the sample mean is

$$\begin{aligned} Var\left(\frac{Y_1 + Y_2 + \ldots + Y_n}{n}\right) &= \frac{1}{n^2}Var(Y_1) + \frac{1}{n^2}Var(Y_2) + \ldots + \frac{1}{n^2}Var(Y_n) \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Summary

- Looked at continuous random variables
  - There are differences, but much remains the same
- Looked at combination of random variables
  - Expectation
  - Variance
- Next lecture: start developing models for data