The University of Sydney

MATH1005 Statistics

maths.usyd.edu.au/u/UG/WS/WS1005/

Summer/Winter/Semester2 **Tutorial Solutions 8** 2015

This tutorial explores linear functions and sums of random variables.

Linear Function of a Random Variable

Given a random variable X with $E(X)$ and $Var(X)$.
For constants a and b,
$E(a+bX) = a + bE(X)$ and $Var(a+bX) = b^2 Var(X)$.

Sums of Random Variables

Given a sequence of random variables X_i with $E(X_i)$ and $Var(X_i)$ for $i = 1, \ldots, n$.

$$E\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} E(X_i).$$

If X_i are independent, $Var\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} Var(X_i)$.

Sums of Normal Random Variables

Given a sequence of independent random variables $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, n$.

$$\sum_{i=1}^{n} a_i X_i \sim N\Big(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\Big)$$

Given a sequence of iid random variables $X_i \sim N(\mu, \sigma^2)$ and constants a_i for $i = 1, \ldots, n$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\Big(\mu, \frac{\sigma^2}{n}\Big)$$

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

1. Linear Function of a Discrete Random Variable

(a) Consider the following probability distribution function of X.

x	1	2	3	4	Total
$P(X=x)$	0.1	0.2	0.3	0.4	1

Find $E(X)$ and $Var(X)$.

(b) Let $Y = 1 + 2X$. Complete the following probability distribution function of Y.

y	3				Total
$P(Y=y)$	0.1				1

Find $E(Y)$ and $Var(Y)$.

(c) Confirm your answers in (b), by using the formulae for the linear function of a random variable.

Solution

(a) $E(X) = 1 \times 0.1 + 2 \times 0.2 + 3 \times 0.3 + 4 \times 0.4 = 3$
$E(X^2) = 1^2 \times 0.1 + 2^2 \times 0.2 + 3^2 \times 0.3 + 4^2 \times 0.4 = 10$
$Var(X) = 10 - 3^2 = 1.$

(b)

y	3	5	7	9	Total
$P(Y = y)$	0.1	0.2	0.3	0.4	1

$E(Y) = 3 \times 0.1 + 5 \times 0.2 + 7 \times 0.3 + 9 \times 0.4 = 7$
$E(Y^2) = 3^2 \times 0.1 + 5^2 \times 0.2 + 7^2 \times 0.3 + 9^2 \times 0.4 = 53$
$Var(Y) = 53 - 7^2 = 4.$

(c) $E(Y) = E(2X + 1) = 2E(X) + 1 = 2 \times 3 + 1 = 7.$
$Var(Y) = Var(2X + 1) = 4Var(X) = 4 \times 1 = 4.$

2. Sums of Random Variables

Given two random variables X: $E(X) = 5$ and $Var(X) = 9$ and Y: $E(Y) = 3$ and $Var(Y) = 4$.

(a) Find $E(X + Y)$ and $Var(X + Y)$. What condition is necessary to calculate the variance?

(b) Find $E(2X + Y + 3)$ and $Var(2X + Y + 3)$.

(c) Find $E(-5X - 6Y)$ and $Var(-5X - 6Y)$.

Solution

(a) $E(X + Y) = E(X) + E(Y) = 5 + 3 = 8$
$Var(X + Y) = Var(X) + Var(Y) = 9 + 4 = 13$ (assuming independence)

Condition: Independence of X and Y.

(b) $E(2X + Y + 3) = 2E(X) + E(Y) + 3 = 2 \times 5 + 3 + 3 = 16$
$Var(2X + Y + 3) = 4Var(X) + Var(Y) = 4 \times 9 + 4 = 40$

(c) $E(-5X - 6Y) = -5E(X) - 6E(Y) = -5 \times 5 - 6 \times 3 = -43$
$Var(-5X - 6Y) = (-5)^2Var(X) + (-6)^2Var(Y) = 25 \times 9 + 36 \times 4 = 369$

3. Sample Mean and Total of iid Normal Random Variables

Suppose $X_i \sim N(100, 125)$ for $i = 1, \ldots, 5$.

(a) Find the distribution of $\bar{X}$ and $T = \sum_{i=1}^{5} X_i$.

(b) Calculate $P(95 \leq \bar{X} \leq 105)$. Is this what you expected?

(c) $P(T \geq 600)$. Is this what you expected?

Solution

Given: $X_i \sim N(\mu = 100, \sigma^2 = 125)$ for $i = 1, \ldots, n$, where $n = 5$.

(a) $\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) = N(100, \frac{125}{5}) = N(100, 25) = N(100, 5^2).$

$T = \sum_{i=1}^{5} X_i \sim N(n\mu, n\sigma^2) = N(5 \times 100, 5 \times 125) = N(500, 625) = N(500, 25^2)$.

(b) Standardising we get

$$P(95 \leq \bar{X} \leq 105) = P(\frac{95-100}{5} \leq \frac{\bar{X}-100}{5} \leq \frac{105-100}{5}) = P(-1 \leq Z \leq 1)$$

which is

$$P(Z \leq 1) - P(Z \leq -1) = \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1 \approx 0.68$$

This is what we expect: as 68% of probability lies within 1 standard deviation of the mean.

(c) Standardising we get

$$P(T \geq 600) = P(\frac{T-500}{25} \geq \frac{600-500}{25}) = P(Z \geq 4) \approx 0$$

This is what we expect: as it would be highly unlikely to observe a value more than 4 sds away from the mean.

4. Mean of iid Normal Random Variables

An electrical firm manufactures light bulbs. The lifetime of the bulbs is approximately normally distributed, with average lifetime of 800 hours and variance of 1600 hours.

(a) Find the probability that a randomly chosen light bulb lasts less than 790 hours.

(b) Find the probability that the average lifetime of a sample of 25 bulbs is less than 790 hours.

Solution

X = lifetime of bulb $\sim N(800, 1600)$.

(a) $P(X < 790) = P(\frac{X-800}{40} < \frac{790-800}{40}) = P(Z < -1/4) = P(Z > 1/4) = 1 - \Phi(1/4) = 0.4013$

Check in R:

```
> pnorm(790,800,40)
[1] 0.4012937
```

(b) $\bar{X}$ = average lifetime of bulb $\sim N(\mu, \sigma^2/n) = N(800, 1600/25) = N(800, 64)$.

$$P(\bar{X} < 790) = P(\frac{\bar{X}-800}{8} < \frac{790-800}{8}) = P(Z < -10/8) = P(Z > 1.25) = 1 - \Phi(1.25) = 0.1056$$

Check in R:

```
> pnorm(790,800,8)
[1] 0.1056498
```

5. Sum of iid Normal Random Variables

Suppose that the weight (in pounds) of a North American adult can be represented by a normal random variable with mean 150 and variance 625. An elevator contains a sign "Maximum 10 people". It can in fact safely carry 2000 lbs.

(a) (Extension) If we want to be at least 99% certain that the elevator will not be overloaded, what is the maximum number of people that should be allowed into the elevator?

(b) If we want an elevator that we are 99% certain can carry 10 people without overloading, how much weight should it be able to safely carry?

Solution

X_i = weight of a North American adult $\sim N(150, 625)$, $i = 1, 2, \ldots, n$

Safety limit for the lift:
Sign says: 'Maximum 10 people'
Actual limit: 2000 pounds

(a) We want to find n such that $P(\text{total weight} \leq \text{safety limit}) = P(\sum_{i=1}^{n} X_i \leq 2000) \geq 0.99$.

Now $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2) = N(150n, 625n)$.

So $P(\sum_{i=1}^{n} X_i \leq 2000) = P(\frac{\sum_{i=1}^{n} X_i - 150n}{25\sqrt{n}} \leq \frac{2000 - 150n}{25\sqrt{n}}) = P(Z \leq \frac{2000 - 150n}{25\sqrt{n}})$

Hence we want to find n such that $P(Z \leq \frac{2000 - 150n}{25\sqrt{n}}) \geq 0.99$.

From the Normal table, the 99% quantile of Z is 2.33.

Hence we need to solve $\frac{2000 - 150n}{25\sqrt{n}} \geq 2.33$.

(1) Method1: Using quadratic
Rearranging, we get $150n + 58.25\sqrt{n} - 2000 = 0$, which is a quadratic in $\sqrt{n}$.

Using the quadratic formula,
$\sqrt{n} = \frac{-58.25 \pm \sqrt{58.25^2 - 4 \times 150 \times -2000}}{2 \times 150}$
so $\sqrt{n} = -3.850809$ or 3.462476, giving $n = (3.462476)^2 = 11.98874$.
Hence, we would choose $n = 11$ to be safe.

(2) Method2: Using trial and error

n	$\frac{2000 - 150n}{25\sqrt{n}}$
9	8.7
10	6.3
11	4.2
12	2.3
13	0.6

To get $\frac{2000 - 150n}{25\sqrt{n}} \geq 2.33$, we need a maximum of $n = 11$.

(b) We want to find L such that $P(\sum_{i=1}^{10} X_i \leq L) \geq 0.99$.

Now $\sum_{i=1}^{10} X_i \sim N(1500, 6250)$.

So $P(\sum_{i=1}^{10} X_i \leq L) = P(\frac{\sum_{i=1}^{10} X_i - 1500}{\sqrt{6250}} \leq \frac{L - 1500}{\sqrt{6250}}) = P(Z \leq \frac{L - 1500}{\sqrt{6250}})$

Hence we want $P(Z \leq \frac{L - 1500}{\sqrt{6250}}) \geq 0.99$.

Solving $2.33 = \frac{L - 1500}{\sqrt{6250}}$, we get $L = 2.33 * \sqrt{6250} + 1500 = 1684.203$.
Hence, choose a safety limit of 1684 pounds.

6. A pharmaceutical firm produces a rare antibiotic. It is believed that daily production (in mg) of the drug may be represented by a normal random variable with mean 500 and variance 1600.

(a) A hospital wants as much of the antibiotic as possible five weeks from now. Assuming independence of the production each day and 5 working days per week, how much of the antibiotic can the firm promise, and be 90% certain of keeping its promise?

(b) (Extension) After delivering the antibiotic to the hospital, the firm receives an urgent order for 10 grams of the drug from a hospital in Iqaluit. How quickly can the hospital promise to deliver the order, and be 95% certain of keeping its promise?

(c) After studying its production more closely, the firm realizes that its production (in mg/day) varies due to absenteeism. The daily means and variances are as follows:

Day	Mean	Variance
Monday	300	900
Tuesday	500	1000
Wednesday	600	1100
Thursday	500	1000
Friday	300	900

If it can be assumed that each day's production is normally distributed and independent of the other days, what is the probability that the total production in a week exceeds 2 grams?

(d) Verify your answers in (a) and (c) by using the R commands `pnorm` and `qnorm`.

Solution

Daily Production $= X_i \sim N(500, 1600), i = 1, 2, 3, \dots$

(a) Total production in 25 days $= \sum_{i=1}^{25} X_i \sim N(25 \times 500, 25 \times 1600) = N(12500, 200^2)$

We want L such that $P(\sum_{i=1}^{25} X_i \geq L) \geq 0.90$.

Hence $P(\frac{\sum_{i=1}^{25} X_i - 12500}{200} \geq \frac{L - 12500}{200}) \geq 0.90$.

Solving $-1.28 = \frac{L - 12500}{200}$, we get $L = -1.28 \times 200 + 12500 = 12244$.
Hence, the firm could promise a production of 12244g.

(b) We want n (number of days) such that $P(\sum_{i=1}^{n} X_i \geq 10000) \geq 0.95$.

Total production in n days $= \sum_{i=1}^{n} X_i \sim N(500n, 1600n)$

So we want $P(\sum_{i=1}^{n} X_i \geq 10000) = P(\frac{\sum_{i=1}^{n} X_i - 500n}{\sqrt{1600n}} \geq \frac{10000 - 500n}{\sqrt{1600n}}) = P(Z \geq \frac{10000 - 500n}{\sqrt{1600n}}) \geq 0.95$

We need to solve $-1.645 = \frac{10000 - 500n}{\sqrt{1600n}}$.

(1) Method1: Using quadratic
Rearranging gives the quadratic in $\sqrt{n}$: $500n - 65.8\sqrt{n} - 10000 = 0$.

Using the quadratic formula, we get $\sqrt{n} = \frac{65.8 \pm \sqrt{-65.8^2 - 4 \times 500 \times -10000}}{2 \times 500}$
which gives $\sqrt{n} = -4.40682$ and 4.53842.

Hence $n = 21$. The hospital expects to deliver within 21 days.

(2) Method2: Using trial and error

n	$\dfrac{10000-500n}{\sqrt{1600n}}$
19	2.867697
20	0
21	-2.727724
22	-5.330018

To get $\dfrac{10000-500n}{\sqrt{1600n}} \geq 1.645$, we need a maximum of $n = 21$.

(c) Total Production in a week = $T \sim N(300+500+600+500+300, 900+1000+1100+1100+900)$.
So $T \sim N(2200, 5000)$.
Hence $P(T \geq 2000) = P(\dfrac{T-2200}{\sqrt{5000}} \geq \dfrac{2000=2200}{\sqrt{5000}}) = P(Z \geq -2.828427) \doteq 0.998$

(d)

```
> qnorm(0.1,12500,200)
[1] 12243.69
> 1-pnorm(2000,2200,70)
[1] 0.9978626
```