STAT 111

Recitation 3

Mo Huang

Email: mohuang@wharton.upenn.edu

Office Hours: Wednesdays 3:00 - 4:00 pm, JMHH F96

Slides (adapted from Gemma Moran): `github.com/mohuangx/STAT111-Fall2018`

September 21, 2018

Random Variables: Mean

- The mean of a random variable (expected value) is the long-run average of realizations of the random variable over repeated experiments.

- This mean of random variable X is different from the sample average, which is the average of a *finite* number of observations ($\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$, where $x_1, \ldots, x_n$ are observed data).

- Let X be a discrete random variable that can take values $\{v_1, v_2, \ldots, v_k\}$. Then the mean of X is given by:

$$\begin{aligned} \mu &= \sum_{i=1}^{k} v_i P(X = v_i) \\ &= v_1 P(X = v_1) + v_2 P(X = v_2) + \cdots + v_k P(X = v_k). \end{aligned}$$

- If X is binomial with parameters n and θ, then the mean of X is

$$\mu = n\theta$$

Random Variables: Variance

- The variance of a random variable is a measure of the *spread* of a distribution - that is, how far away values are from the mean.

- Let X be a discrete random variable that can take values $\{v_1, v_2, \ldots, v_k\}$. Then the variance of X is given by:

$$\sigma^2 = (v_1 - \mu)^2 P(X = v_1) + \cdots + (v_k - \mu)^2 P(X = v_k) \tag{1}$$

or

$$\sigma^2 = v_1^2 P(X = v_1) + \cdots + v_k^2 P(X = v_k) - \mu^2 \tag{2}$$

- The standard deviation of a random variable is the square root of the variance and is denoted by σ.

- For a binomial random variable $X \sim \mathcal{B}(n, \theta)$:

$$\sigma^2 = n\theta(1 - \theta)$$

Random Variables: Questions

Q1: Let X be a random variable with the below distribution. Find the mean and the variance of X using formula (1) and then formula (2).

x	-3	-1	4	5
$P(X = x)$	0.1	0.3	0.4	0.2

Table: Probability distribution of X.

A1: $\mu = -3 \times 0.1 - 1 \times 0.3 + 4 \times 0.4 + 5 \times 0.2 = 2.$

$$\sigma^2 = (-3-2)^2(0.1) + (-1-2)^2(0.3) + (4-2)^2(0.4) + (5-2)^2(0.2)$$
$$= 8.6.$$

$$\sigma^2 = (-3)^2(0.1) + (-1)^2(0.3) + 4^2(0.4) + 5^2(0.2) - 2^2$$
$$= 8.6.$$

Random Variables: Questions

Q2: Let X be a random variable with the below distribution. Write the probability distribution of $Y = 2X$ in tableau form.

x	-3	-1	4	5
$P(X = x)$	0.1	0.3	0.4	0.2

Table: Probability distribution of X.

A2: Probability distribution of Y:

y	-6	-2	8	10
$P(Y = y)$	0.1	0.3	0.4	0.2

Table: Probability distribution of Y.

Properties of mean and variance

- Suppose X is a random variable with mean μ and variance σ^2. Consider fixed number c.

- Mean properties
  - Mean of $X + c \Longrightarrow \mu + c$.
  - Mean of $cX \Longrightarrow c\mu$.

- Variance properties
  - Variance of $X + c \Longrightarrow \sigma^2$.
  - Variance of $cX \Longrightarrow c^2\sigma^2$.

Q3: Suppose X has mean $\mu_X = 3$ and variance $\sigma_X^2 = 4$. What is the mean and variance of $Y = 2 + 5X$?

A3: $\mu_Y = 2 + 5\mu_X = 2 + 5(3) = 17$

$\sigma_Y^2 = 5^2\sigma_X^2 = 5^2(4) = 100$

Many Random Variables

- Suppose we plan an experiment with a sample size of n.
- We have n future outcomes, or random variables, denoted by: $X_1, X_2, \ldots, X_n$.
- We say $\{X_1, \ldots, X_n\}$ are independently and identically distributed (or i.i.d) if:
  - All the X_is are independent of each other.
  - Each X_i has the same probability distribution.
- For example:
  - I plan to roll a dice n times. X_i represents the future outcome of the ith roll. X_i is *independent* of the other rolls of the dice, and it has the same probability of getting a $1, 2, 3, 4, 5, 6$ as the other rolls.

Many Random Variables

- Recall the sample average is given by:

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}.$$

- What if we haven't observed the data $x_1, \ldots, x_n$ yet?
- Before an experiment, the average is *also* a random variable:

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

- The *sum* is also a random variable:

$$T_n = X_1 + \cdots + X_n$$

Many Random Variables

- Let X and Y be two random variables. Then,

  $$\text{mean}(X + Y) = \text{mean}(X) + \text{mean}(Y).$$

  - If X, Y are also independent,

    $$\text{variance}(X + Y) = \text{variance}(X) + \text{variance}(Y).$$

  - For constants a, b, we have

    $$\text{mean}(aX + bY) = a \times \text{mean}(X) + b \times \text{mean}(Y)$$

    $$\text{variance}(aX + bY) = a^2 \times \text{variance}(X) + b^2 \times \text{variance}(Y).$$

- Let $D = X - Y$. What is the variance of D?

$$\text{variance}(D) = \text{variance}(X) + \text{variance}(Y).$$

Many Random Variables

- Let $X_1, \ldots, X_n$ be i.i.d. random variables, each with mean μ and variance σ^2.
- Then for the sum, T_n:

$$\text{mean of } T_n = n\mu, \quad \text{variance of } T_n = n\sigma^2$$

- For the average, $\bar{X}$:

$$\text{mean of } \bar{X} = \mu, \quad \text{variance of } \bar{X} = \frac{\sigma^2}{n}.$$

Questions

Q4: Suppose we have a company producing a medicine. Each day the *mean* amount of medicine produced is 500 mg and the *variance* is 900 mg^2. Assume the amount produced each day is independent.

(i) Let T_n be the total amount of medicine produced in a week. Find the mean and variance of T_n.

$Mean(T_n) = 5 \times 500 = 2500$
$Var(T_n) = 5 \times 900 = 4500$

(ii) Let $\bar{X}$ be the average amount of medicine produced in a 5-day week. Find the mean and variance of $\bar{X}$.

$Mean(\bar{X}) = 500$
$Var(\bar{X}) = 900/5 = 180$

Proportions

- Sometimes it is necessary to consider the *proportion* of "successes" in a Binomial trial (instead of the total number of successes)

- Proportions are a type of average!

- Let

$$Y_i = \begin{cases} 1 & \text{if the } i\text{th trial is a "success"} \\ 0 & \text{if the } i\text{th trial is a "failure"} \end{cases}$$

- Then we have $Y_1, \ldots, Y_n$ where $Y_i \sim \mathit{Binomial}(1, \theta)$.

- The proportion of successes is the average:

$$P = \frac{Y_1 + \cdots + Y_n}{n}$$

Proportions

- $Y_i \sim Binomial(1, \theta)$. Recall:

$$Mean(Y_i) = \theta, \quad Var(Y_i) = \theta(1-\theta).$$

- The proportion of successes is the average:

$$P = \frac{Y_1 + \cdots + Y_n}{n}$$

- Then,

$$Mean(P) = \theta, \quad Var(P) = \frac{\theta(1-\theta)}{n}$$

.

Q5: Suppose we plan to toss a coin 20 times and $Prob(H) = 0.7$. Let P be the *proportion* of heads that we toss. Find the mean and variance of P.

A5:

$$Mean(P) = 0.7, \quad Var(P) = 0.7 \times 0.3/20 = 0.0105.$$

Questions

Q6: Suppose the company producing a medicine has different means and variances the amount produced on each day of the week:

Day	Mean	Variance
Monday (X_1)	450	1200
Tuesday (X_2)	550	800
Wednesday (X_3)	600	500
Thursday (X_4)	550	800
Friday (X_5)	350	1200

Find mean and variance of both the sum T_n and the average $\bar{X}$.

A6: X_1, X_2, X_3, X_4, and X_5 are no longer $i.i.d.$!

$$Mean(T_n) = 450 + 550 + 600 + 550 + 350 = 2500$$
$$Var(T_n) = 1200 + 800 + 500 + 800 + 1200 = 4500$$

$$Mean(\bar{X}) = 1/n \times Mean(T_n) = 500$$
$$Var(\bar{X}) = 1/n^2 \times Var(T_n) = 4500/25 = 180$$

Questions

Q7: Let P_1 be the proportion of heads in 50 coin tosses, where $P(H) = 0.6$. Find $Mean(P_1)$ and $Var(P_1)$.

A7: $Mean(P_1) = 0.6$ and $Var(P_1) = 0.6 \times 0.4/50 = 0.0048$.

Q8: Let P_2 be the proportion of heads in 20 coin tosses, where $P(H) = 0.7$. From earlier, $Mean(P_2) = 0.7$ and $Var(P_2) = 0.0105$. Let $D = P_1 - P_2$. Find the mean and variance of D.

A8: $Mean(D) = 0.6 - 0.7 = -0.1$
$Var(D) = 0.0048 + 0.0105 = 0.0153$

Continuous Random Variables

- So far, we have just considered discrete random variables; those whose possible values are countable.
- A continuous random variable can take continuous values in a future experiment.
- Every continuous random variable X has an associated density function $f(x)$.

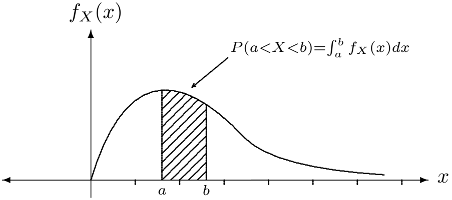
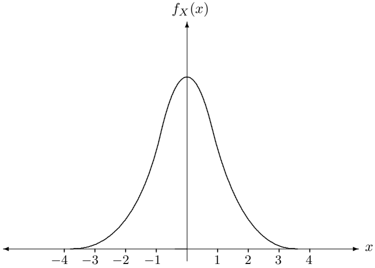


Figure 5: $P(a < X < b) = \int_a^b f_X(x)dx$.

The Normal Distribution

- A normal random variable is a continuous random variable.



The density function for the standard normal distribution with $\mu = 0$, $\sigma = 1$.

- We call a normal random variable with $\mu = 0$ and $\sigma^2 = 1$ a standard normal random variable.

- For standard normal random variables, we can use charts (or a computer) to find the area under the density function (i.e. the probabilities).

Questions

- $P(Z < -1.75)$

  $P(Z < -1.75) = 0.0401$

- $P(Z > 0.85)$

  $P(Z > 0.85) = 1 - 0.8023 = 0.1977$

- $P(-1.43 < Z < 0.92)$

  $P(-1.43 < Z < 0.92) = 0.8212 - .0764 = .7448$