Discrete Random Variables

Chapter 3 – Lecture 11

Yiren Ding

Shanghai Qibao Dwight High School

March 31, 2016

Theorem 1 (Linearity Property).

For any two discrete random variables X, Y defined on the **same** *sample space,*

$$E(X+Y) = E(X) + E(Y),$$

provided that both $E(X)$ and $E(Y)$ exist and are finite.

- *Proof.* Let X and Y be two random variables defined on the same sample space with probability measure P.
- For example, for the experiment of rolling two dice, X is the smallest of the two outcomes, and Y is the sum of the two outcomes.
- We define a new random variable $Z = X + Y$, with the following probability mass function

$$P(Z = z) = \sum_{x+y=z} P(X = x, Y = y),$$

where $P(X = x, Y = y)$ denotes the probability of the joint event that X takes on the value x and Y the value y.

Theorem 1

- By definition, $E(Z) = \sum_z zP(Z = z)$, so we have

$$\begin{aligned}
E(Z) &= \sum_z \sum_{x+y=z} (x+y)P(X = x, Y = y) \\
&= \sum_{x,y} (x+y)P(X = x, Y = y) \\
&= \sum_{x,y} xP(X = x, Y = y) + \sum_{x,y} yP(X = x, Y = y) \\
&= \sum_x x \sum_y P(X = x, Y = y) + \sum_y y \sum_x P(X = x, Y = y) \\
&= \sum_x xP(X = x) + \sum_y yP(Y = y) \quad \text{(Rule 1)} \\
&= E(X) + E(Y) \qquad \square
\end{aligned}$$

Indicator Random Variable

- Theorem 1 also works for a finite number of random variables.
- If $E(X_i)$ exists and is finite for all $i = 1, ..., n$, then a repeated application of Theorem 1 gives

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n).$$

- This is extremely useful because it does not require the X_i's to be independent, but only requires them to exist.
- Next, we present a powerful method of finding the expected value of a random variable by representing it as the sum of so-called **indicator random variables** whose range consists of only values 0 and 1.

Example 1 (continue in Example 5).

Suppose that n children of differing heights are placed in line at random. You then select the first child from the line and walk with her/him along the line until you encounter a child who is taller or until you have reached the end of the line. If you do encounter a taller child, you also have her/him to accompany you further along the line until you encounter yet again a taller child or reach the end of the line, etc. Let the random variable X denote the number of children to be selected from the line. What is the expected value of X?

- We can compute $E(X)$ by writing

$$X = X_1 + \cdots + X_n,$$

where the **indicator variable** X_i is defined as

$$X_i = \begin{cases} 1 & \text{if the } i\text{th child is selected from the line} \\ 0 & \text{otherwise.} \end{cases}$$

Example 1 solution

- The probability that the *i*th child is the tallest among the first *i* children is $P(X_i = 1) = \frac{(i-1)!}{i!} = \frac{1}{i}$. Hence,

$$E(X_i) = 0 \times \left(1 - \frac{1}{i}\right) + 1 \times \frac{1}{i} = \frac{1}{i}, \quad i = 1, ..., n$$

- This gives

$$E(X) = \sum_{i=1}^{n} X_i = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$$

- This is the harmonic sum again! Using the *Harmonic Approximation Theorem*, we find that this is probability can be approximated by $\ln n + \gamma$ where $\gamma = 0.57722...$ is the Euler-Mascheroni constant.

Theorem 2 (Substitution Rule).

For any function g of the random variable X,

$$E(g(X)) = \sum_{x \in I} g(x) P(X = x)$$

provided that $\sum_{x \in I} |g(x)| P(X = x) < \infty$.

- *Proof.* If the random variable X takes on the values $x_1, x_2, \ldots$ with probabilities $p_1, p_2, \ldots$ and assume $g(x_i) \neq g(x_j)$ for $x_i \neq x_j$.
- Then the random variable $Z = g(X)$ takes on values $z_1 = g(x_1)$, $z_2 = g(x_2), \ldots$ with the same probabilities $p_1, p_2, \ldots$.
- Therefore $E(Z) = \sum_k z_k P(Z = z_k) = \sum_k g(x_k) P(X = x_k)$.
- Obviously the proof needs a slight modification if g is not one-to-one, i.e., it is possible to have $x_i \neq x_j$ such that $g(x_i) = g(x_j)$.

Theorem 3.

For any random variable X and constants a and b,

$$E(aX + b) = aE(X) + b.$$

- The proof is trivial and is left as a homework exercise.
- Warning: In general $E(g(X)) \neq g(E(X))$, or stated differntly, the average value of the input X does not determine in general the average value of the output $g(X)$!
- Consider a simple counterexample by taking the random variable X with $P(X = 1) = P(X = -1) = \frac{1}{2}$ and take the function $g(x) = x^2$. In this case $E(g(X)) = 1$ whereas $g(E(X)) = 0$.
- This theorem together with linearity will be very useful to us.

Definition 1 (Variance).

If X is a random variable, and let $\mu = E(X)$. Then the **variance** of X is simply the expected value of $(X-\mu)^2$, denoted by

$$\operatorname{var}(X) = E((X-\mu)^2).$$

- Variance measures the spread of the values of X around the mean.
- Often one uses the **standard deviation** defined by

$$\sigma(X) = \sqrt{\operatorname{var}(X)}$$

  since it has the same units as $E(X)$.
- Another formula for variance can be derived by expanding

$$\begin{aligned}\operatorname{var}(X) &= E((X-\mu)^2) = E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 = \boxed{E(X^2) - \mu^2}\end{aligned}$$

Example 2.

For any constants a and b, prove that

$$\text{var}(aX + b) = a^2\text{var}(X).$$

- Let $\mu = E(X)$, and by Theorem 2, $E(aX + b) = a\mu + b$. We have

$$\begin{aligned}
\text{var}(aX + b) &= E((aX + b)^2) - (a\mu + b)^2 \\
&= E(a^2X^2 + 2abX + b^2) - (a\mu + b)^2 \\
&= a^2E(X^2) + \cancel{2ab\mu} + \cancel{b^2} - a^2\mu^2 - \cancel{2ab\mu} - \cancel{b^2} \\
&= a^2(E(X^2) - \mu^2) \\
&= a^2\text{var}(X)
\end{aligned}$$

- From this it follows that

$$\sigma(aX + b) = a\sigma(X).$$

Example 3.

What is the variance of the total score of a roll of two dice?

- Let the random variable X denote the total score. X takes on values $2, 3, \ldots, 12$ with a triangular distribution.
- The expected value of X is

$$E(X) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} + 7 \times \frac{6}{36}$$
$$+8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} = 7$$

- The variance is

$$\text{var}(X) = 2^2 \times \frac{1}{36} + 3^2 \times \frac{2}{36} + 4^2 \times \frac{3}{36} + 5^2 \times \frac{4}{36} + 6^2 \times \frac{5}{36} + 7^2 \times \frac{6}{36}$$
$$+8^2 \times \frac{5}{36} + 9^2 \times \frac{4}{36} + 10^2 \times \frac{3}{36} + 11^2 \times \frac{2}{36} + 12^2 \times \frac{1}{36} - 7^2 = 5\frac{5}{6}$$

Example 4.

Suppose the random variable X has the Poisson distribution $P(X = k) = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, \ldots$. What are $E(X)$ and $\mathrm{var}(X)$?

- We will learn more about this distribution next week. For now, let's just verify the remarkable fact that if X is Poisson, then

$$\mathrm{var}(X) = E(X) = \lambda.$$

- To show that we use the definition $E(X) = \sum_{n=0}^{\infty} nP(X = n)$,

$$\begin{aligned} E(X) &= \lambda e^{-\lambda} + 2\frac{\lambda^2}{2!}e^{-\lambda} + 3\frac{\lambda^3}{3!}e^{-\lambda} + \cdots \\ &= \lambda e^{-\lambda}\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots\right) = \lambda e^{-\lambda}e^{\lambda} = \lambda. \end{aligned}$$

- The last step uses the well-known power series $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots$

Example 4 solution

- To find a variance, let's use a clever method.
- Recall that $\text{var}(X) = E(X^2) - (E(X))^2$.
- To find $E(X^2)$ we use the identity $k^2 = k(k-1) + k$. This gives

$$E(X^2) = \sum_{k=0}^{\infty} k^2 P(X = k) = \sum_{k=1}^{\infty} k(k-1)P(X = k) + \sum_{k=1}^{\infty} kP(X = k)$$
$$= \sum_{k=1}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} + E(X) = \lambda^2 \sum_{k=2}^{\infty} e^{-\lambda}\frac{\lambda^{k-2}}{(k-2)!} + \lambda.$$

- Since $\sum_{k=2}^{\infty} e^{-\lambda}\lambda^{k-2}/(k-2)! = \sum_{n=0}^{\infty} e^{-\lambda}\lambda^n/n! = 1$, we have $E(X^2) = \lambda^2 + \lambda$. And it follows that $\text{var}(X) = \lambda$.

Example 5 (Example 1 cont'd).

Continuing from Example 1, what is the variance of the random variable X denoting the number of children selected from the line?

- Again we use the representation $X = X_1 + \cdots + X_n$, where X_i is the indicator variable equal to 1 if the ith child is the tallest among the first $i-1$ children and equal to 0 otherwise.
- Then we have

$$E(X^2) = \sum_{i=1}^{n} E(X_i^2) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} E(X_iX_j).$$

- Recall that $P(X_i = 1) = \frac{1}{i}$, it follows that $E(X_i^2) = \frac{1}{i}$ for all i.
- To find $E(X_iX_j)$ for $i \neq j$, note that $X_iX_j = 1$ only if $X_i = X_j = 1$ and is 0 otherwise. Hence, $E(X_iX_j) = P(X_i = 1, X_j = 1)$.

Example 5 solution

- For any $i < j$, the joint probability is

$$P(X_i = 1, X_j = 1) = \frac{(i-1)! \times 1 \times (i+1) \times \cdots \times (j-1) \times 1}{j!} = \frac{1}{i} \times \frac{1}{j}.$$

- Therefore,

$$E(X^2) = \sum_{i=1}^{n} \frac{1}{i} + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{1}{i} \times \frac{1}{j}.$$

- Since $E(X) = \sum_{i=1}^{n} \frac{1}{i}$, we have $(E(X))^2 = \sum_{i=1}^{n} \frac{1}{i^2} + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{1}{i} \times \frac{1}{j}$.
- Since $\text{var}(X) = E(X^2) - (E(X))^2$, we find that

$$\text{var}(X) = \sum_{i=1}^{n} \frac{1}{i} - \sum_{i=1}^{n} \frac{1}{i^2} \approx \ln n + \gamma - \frac{\pi^2}{6}.$$

Example 6 (Newsboy Problem).

Every morning, rain or shine, newspaper girl Violet can be found at the entrance to the metro, hawking copies of "The Shanghai Daily." Demand for newspapers varies from day to day, but Violet's regular early morning haul yields her 200 copies. She purchases these copies for 1 RMB per paper, and sells them for 1.50 RMB apiece. Violet goes home at the end of the morning, or earlier if she sells out. She can return unsold papers to the distributor for 0.50 RMB apiece. From experience, Violet knows that demand for papers on any given morning is uniformly distributed between 150 and 250, where each of the possible values $150, \ldots, 250$ is equally likely. What are the expected value and the standard deviation of Violet's net earnings on any given morning?

- Denote by the random variable X the number of copies Violet could have sold on a given morning if she had unlimited supply.
- The probability mass function of X is $P(X = k) = \frac{1}{101}$ for $k = 150, \ldots, 250$.

Example 6 solution

- Violet's net earnings is a random variable $g(X)$, where

$$g(x) = \begin{cases} -200 + 1.5x + 0.5(200 - x), & x \leq 200 \\ -200 + 1.5 \times 200, & x > 200. \end{cases}$$

- By the substitution rule, we find that Violet's average earning is

$$\begin{aligned} E(g(X)) &= \sum_{k=150}^{250} g(k)P(X = k) \\ &= \frac{1}{101} \sum_{k=150}^{200} (-100 + k) + \frac{1}{101} \sum_{k=201}^{250} 100 \\ &= \frac{3825}{101} + \frac{5000}{101} = 87.3762. \end{aligned}$$

Example 6 solution

- To find the standard deviation of $g(X)$, we first find

$$E((g(X))^2) = \frac{1}{101}\sum_{k=150}^{200}(-100+k)^2 + \frac{1}{101}\sum_{k=201}^{250}100^2$$
$$= \frac{297,925}{101} + \frac{500,000}{101} = 7900.2475$$

- Hence the variance of Violet's net earnings is

$$\mathrm{var}(g(X)) = E((g(X))^2) - (E(g(X)))^2$$
$$= 7900.2475 - (87.3762)^2 = 265.647.$$

- Hence Violet's net earnings has an expected value of 87.38 RMB and a standard deviation of $\sqrt{265.64} = 16.30$ RMB.