Session 82: Variance

- Variance
- Examples

Variance

Definition 4: Let X be a random variable on the sample space S. The **variance** of X, denoted by $V(X)$ is

$$V(X) = \sum_{s \in S} (X(s) - E(X))^2 p(s)$$

The **standard deviation** of *X*, denoted by $\sigma(X)$ is defined as $\sqrt{V(X)}$

- Variance and standard deviation are used to quantify how widely a random variable is distributed

Example

Let X and Y be random variables on S = {1, 2, 3, 4, 5, 6}

Let X(s) = 0 for all $s \in S$

Let Y(s) = -1 for $s \in \{1,2,3\}$ and Y(s) = 1 for $s \in \{4,5,6\}$

$$E(X) = E(Y) = 0$$

$$V(X) = \sum_{s \in S} (\underbrace{X(s) - E(X)}_{0})^2 p(s) = 0$$

$$V(Y) = \sum_{s \in S} \underbrace{(X(s) - E(X))^2}_{1} \underbrace{p(s)}_{\frac{1}{6}} = 6 \cdot \frac{1}{6} = 1$$

Characterisation of Variance

Theorem 6: If X is a random variable on a sample space S, then

$$V(X) = E(X^2) - E(X)^2$$

Corollary 1: If *X* is a random variable on a sample space *S* and *E*(*X*) = μ , then

$$V(X) = E((X - \mu)^2)$$

$$V(X) = \sum_{s \in S} \left(X(s) - E(X) \right)^2 p(s) =$$

$$\underbrace{\sum_{s \in S} (X(s))^2 p(s)}_{E(X^2)} - \underbrace{2 \sum_{s \in S} X(s) \cdot p(s) \cdot E(X)}_{-2E(X) \sum_{s \in S} X(s) p(s)} + \underbrace{\sum_{s \in S} E(X)^2 \cdot p(s)}_{E(X)^2 \sum_{s \in S} p(s) = E(X)^2}$$

$$= -2 E(X) \cdot E(X)$$

$$= E(X^2) - E(X)^2$$

$\mu = E(X)$

$E(\mu^2) = \mu^2$

$$E\left((X-\mu)^2\right) = E\left(X^2 - 2\mu X + \mu^2\right) =$$

$$= E(X^2) - 2\mu\, E(X) + \mu^2 =$$

$$= E(X^2) - \mu^2 = E(X^2) - E(X)^2$$

Example

Variance of the Value of a Die: What is the variance of a random variable X, where X is the number that comes up when a fair die is rolled?

We have $V(X) = E(X^2) - E(X)^2$.

We have shown that $E(X) = 7/2$.

We calculate $E(X^2) = 1/6\ (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = 91/6$

and obtain $V(X) = \dfrac{91}{6} - \left(\dfrac{7}{2}\right)^2 = \dfrac{35}{12}$

We used the law of the "unconscious statistician":

$$E(X^2) = \sum_{x \in X(s)} x^2 p(X=x) = \frac{1}{6} \sum_{x \in \{1..6\}} x^2$$

Note: $X(s) \in \{1,2,3,4,5,6\}$
$p(X=x) = \frac{1}{6}, x \in X(s)$

$$E(g(X)) = \sum_{s \in S} g(X(s))\, p(s)$$ Definition

$$E(g(X)) = \sum_{y \in g(X(S))} y\, p(g(X)=y)$$ Theorem for $E(g(X))$

$$E(g(X)) = \sum_{x \in X(s)} g(x)\, p(X=x)$$ different law

Proof: $E(g(X)) = \sum_{s \in S} g(X(s))p(s) = \sum_{x \in X(S)} \sum_{s:X(s)=x} g(X(s))\, p(s) =$

$$= \sum_{x \in X(S)} \sum_{s:X(s)=x} g(x)\, p(s) = \sum_{x \in X(S)} g(x) \sum_{s:X(s)=x} p(s) = \sum_{x \in X(s)} g(x) p(X=x)$$

Variance of Bernoulli Trials

What is the variance of the random variable *X*, where $X(t) = 1$ if a Bernoulli trial is a success and $X(t) = 0$ if it is a failure, where *p* is the probability of success and *q* is the probability of failure?

$$E(X) = p$$

$$V(X) = E(X^2) - E(X)^2 = (p \cdot 1^2 + (1-p) \cdot 0^2) - p^2$$

$$= p - p^2 = p(1-p) = p \cdot q$$

Variance for Independent Random Variables

Bienaymé‘s Formula: If X and Y are two independent random variables on a sample space S, then $V(X + Y) = V(X) + V(Y)$.

Furthermore, if X_i, $i = 1,2, \ldots,n$, with n a positive integer, are pairwise independent random variables on S, then

$$V(X_1 + X_2 + \cdots + X_n) = V(X_1) + V(X_2) + \cdots + V(X_n).$$

Proof : $V(X+Y) = E\left((X+Y)^2\right) - E(X+Y)^2 =$

$$E\left(X^2 + 2XY + Y^2\right) - \left(E(X) + E(Y)\right)^2 =$$

$$E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 =$$

$$\underbrace{E(X^2) - E(X)^2}_{V(X)} + \underbrace{E(Y^2) - E(Y)^2}_{V(Y} + 2\underbrace{E(XY)}_{= E(X) \cdot E(Y) \text{ if } X, Y \text{ independent}} - 2E(X)E(Y) =$$

$$= V(X) + V(Y)$$

Example

Find the variance of the number of successes when n independent Bernoulli trials are performed, where on each trial, p is the probability of success and q is the probability of failure.

Let X_i be the random variable with $X_i((t_1, t_2, \ldots., t_n)) = 1$ if trial t_i is a success and $X_i((t_1, t_2, \ldots., t_n)) = 0$ if it is a failure.

Let $X = X_1 + X_2 + \ldots. X_n$.

Then X counts the number of successes in the n trials.

By Bienaymé 's Formula, it follows that $V(X)= V(X_1) + V(X_2) + \cdots + V(X_n)$.

We have shown that $V(X_i) = pq$ for $i = 1,2, \ldots,n$.

Hence, $V(X) = npq$.

Summary

- Variance
  - Definition
  - Characterisation using expected value
- Examples
  - Bernoulli trials
  - Independent random variables