R. Nazim Khan
Department of Mathematics and Statistics
nazim.khan@uwa.edu.au

The University of Western Australia

3. Probability

Stephen Covey, 7 Habits of Highly Effective People

Habit 2: Begin with the end in mind.

What lies behind us and what lies ahead of us are tiny compared with what lies within us.

Oliver Wendell Holmes

Overview

- We will not repeat the lecture material here, but move forward using that as a foundation.
- We will start with discrete random variables and their properties, and introduce some standard probability distributions.
- We will then consider continuous random variables and their properties. Some standard distributions will be introduced.

3.1 Discrete random variables

- A *random variable* (rv) is mapping from the sample space of a random experiment to the real numbers.
- A *discrete random variable* is one that takes *countably infinite* values.
- The distribution of a discrete rv X is described by a *probability mass function* (pmf), defined as

$$p_X(x) = \mathrm{P}(X = x).$$

- A pdf satisfies

$$p_X(x) \geq 0 \quad \text{and} \quad \sum_{i=1}^{\infty} p_X(x) = 1.$$

3.2 Mean and variance

The *mean* of a random X, denoted $\mathbb{E}(X)$, is defined as

$$\mathbb{E}(X) \sum_{i=-\infty}^{\infty} x \; p_X(x).$$

1. $\mathbb{E}(X)$ is also called the *expectation* or the *mean* of X.
2. We also use the symbol μ_X for $\mathbb{E}(X)$. (μ is the Greek letter mu.)
3. In general, $\mathbb{E}\left(X^2\right) \neq [\mathbb{E}(X)]^2$.

Properties of Mean

E1 $\mathbb{E}(c) = c$ for any constant c.

E2 $\mathbb{E}(aX + b) = a\,\mathbb{E}(X) + b$ for any constants a and b.

E3 $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$. (The mean of a sum is the sum of the means.)

E4. $\mathbb{E}(X - \mu_X) = \mathbb{E}\left[X - \mathbb{E}\left(X\right)\right] = 0$. (Mean shifting.)

Proof

3.3 variance

For a random variable X, we define the *variance* of X, denoted $\mathrm{Var}(X)$, by

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mu_X)^2\right] = \mathbb{E}\left[(X - E(X))^2\right].$$

We also write σ_X^2 for $\mathrm{Var}(X)$.

The *standard deviation* of X, denoted σ_X, is the square root of variance, that is,

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

Properties of Variance

V1. $\text{Var}(X) \geq 0$. (Variance is non-negative.)

V2. $\text{Var}(X) = \mathbb{E}\left(X^2\right) - [\mathbb{E}(X)]^2$. (This is simpler for calculating variance.) (Compare with $\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$.

V3. $\text{Var}(aX + b) = a^2\,\text{Var}(X)$. (Scaling.) (Compare with $u_i = ax_i + b \Rightarrow s_u = a^2\ s_x^2$.)

Proof

V1.

$$\text{Var}(X) = \mathbb{E}\left[(X - \mu_X)^2\right] = \sum_x \underbrace{(x - \mu_X)^2}_{\geq 0}\ \underbrace{p_X(x)}_{\geq 0} \geq 0,$$

as each term in the sum is non-negative (≥ 0).

Proof (ctd)

V2.

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}\left[(X-\mu_X)^2\right] \\
&= \mathbb{E}\left[(X-\mu_X)(X-\mu_X)\right] \\
&= \mathbb{E}\left[X(X-\mu_X)-\mu_X(X-\mu_X)\right] \\
&= \mathbb{E}\left[X^2-\mu_X\,X\right]-\mu_X\,\underbrace{\mathbb{E}(X-\mu_X)}_{=0} \\
&= \mathbb{E}\left(X^2\right)-\mu_X\,\mathbb{E}(X) \\
&= \mathbb{E}\left(X^2\right)-\mu_X^2=\mathbb{E}\left(X^2\right)-\left(\mathbb{E}(X)\right)^2.
\end{aligned}$$

Proof (ctd)

V3. See slide 15. The proof is similar, with Σ replaced by $\mathbb{E}$.

$$
\begin{aligned}
\mathrm{Var}(aX+b) &= \mathbb{E}\left[\left(aX+b-\mathbb{E}\left(aX+b\right)\right)^2\right] \\
&= \mathbb{E}\left[\left(aX+b-a\,\mathbb{E}(X)-b\right)^2\right] \\
&= \mathbb{E}\left(\left[a(X-\mu_X)\right]^2\right) \\
&= \mathbb{E}\left[a^2(X-\mu_X)^2\right] \\
&= a^2\ \mathbb{E}\left[\left(X-\mu_X\right)^2\right] \\
&= a^2\ \mathrm{Var}(X).
\end{aligned}
$$

Note that the standard deviation of $aX+b$, denoted σ_{aX+b}, is $\mid a \mid \sigma_X$.

Standardised random variable

Let the random variable X have mean μ_X and standard deviation σ_X. Put

$$Z = \frac{X - \mu_X}{\sigma_X} = \underbrace{\frac{1}{\sigma_X}}_{a} X - \underbrace{\frac{\mu_X}{\sigma_X}}_{b}.$$

Then

$$\mathbb{E}(Z) = \mathbb{E}(aX - b) = a\,\mathbb{E}(X) - b = \frac{1}{\sigma_X}\,\mu_X - \frac{\mu_X}{\sigma_X} = 0,$$

$$\mathrm{Var}(Z) = \mathrm{Var}(aX - b) = a^2\,\mathrm{Var}(X) = \frac{1}{\sigma_X^2}\,\mathrm{Var}(X) = \frac{1}{\sigma_X^2} \times \sigma_X^2 = 1.$$

We call Z a *standardised random variable*. This result is similar to that for standardising data (see §1.8):

$$z_i = \frac{x_i - \overline{x}}{s} \Rightarrow \overline{z} = 0 \text{ and } s_z^2 = 1.$$

Note

$$
\begin{aligned}
Y &= aX + b \\
\Rightarrow \mathrm{Var}(Y) &= a^2 \mathrm{Var}(X) \\
\text{and } \sigma_Y &= |a| \ \sigma_X,
\end{aligned}
$$

since standard deviation is always non-negative. This is similar to the result for linear transformation of data. Note that

$$
|a| = \begin{cases} a \text{ if } a \geq 0, \\ -a \text{ if } a < 0. \end{cases}
$$

Facts

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$$

$$\sum_{i=1}^{\infty} i x^{i-1} = \frac{1}{(1-x)}^{2}.$$

3.3 Some standard distributions

Bernoulli distribution

A random variable $X \sim \mathit{Bern}(p)$ if

$$p_X(x) = p^x(1-p)^{1-x}, \quad x = 0, 1.$$

Then

$$\mathbb{E}(X) = p, \mathrm{Var}(X) = p(1-p).$$

Geometric distribution

Consider a sequence of Bernoulli trials, and let the random variable$X$ denote the trial on which the first success occurs. Then $X \sim Geom(p)$ where p is the probability of success, with probability mass function

$$p_X(x) = p(1-p)^{x-1},\ quadx = 1, 2, 3, \ldots.$$

Now

$$\mathbb{E}(X) = \frac{1}{p} \text{ and } \mathsf{Var}(X) = \frac{1-p}{p^2}.$$

Proof

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{x=1}^{\infty} xp(1-p)^{x-1} \\
&=
\end{aligned}
$$

Proof (ctd)

Geometric distribution: variance

$$
\begin{aligned}
\mathbb{E}(X^2) &= \sum_{x=1}^{\infty} x^2 p(1-p)^{x-1} \\
&=
\end{aligned}
$$

Geometric distribution: variance (ctd)

Binomial distribution

Let the random variable X denote the number of successes in n iid Bernoulli trials. Then $X \sim Bin(n, p)$ where p is the probability of success, with probability mass function

$$p_X(k) = \binom{n, k}{p}^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$

Then

$$\mathbb{E}(X) = np, \mathrm{Var}(X) = np(1-p).$$

Proof

Proof(ctd)

Proof(ctd)

Poisson distribution

Let the random variable X denote the number of successes in a fixed volume. Then $X \sim Poi(\lambda)$ where λ is the mean number of occurrences, with probability mass function

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

Now

$$\mathbb{E}(X) = \lambda \text{ and } \mathrm{Var}(X) = \lambda.$$

Proof

Proof(ctd)

Proof(ctd)