Statistics of Continuous Random Variables

Expected Value / Expectation

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x$$

Expected Value and the Mean

Since $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$, the expected value of a random variable is equal to the mean of the random variable.

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x = \frac{\int_{-\infty}^{\infty} xf(x)\mathrm{d}x}{\int_{-\infty}^{\infty} f(x)\mathrm{d}x} = \mu$$

Expected Value of a Function of a Random Variable

$$\mathrm{E}[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x$$

Expected Value of Linear Combinations of Random Variables

$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$

$\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$

Example VCAA 2017 NHT Exam 2 Question 3b

A company supplies schools with whiteboard pens. The total length of time for which a whiteboard pen can be used for writing before it stops working is called its use-time. There are two types of whiteboard pens: Grade A and Grade B. The use-time of Grade B whiteboard pens is described by the probability density function f where x is the use-time in hours

$$f(x) = \begin{cases} \frac{x}{576}(12 - x)\left(e^{\frac{x}{6}} - 1\right) & 0 \le x \le 12 \\ 0 & \text{otherwise} \end{cases}$$

The expected use-time of a Grade B whiteboard pen is

$$\mathrm{E}(X) = \int_{0}^{12} xf(x)\mathrm{d}x = \int_{0}^{12} \frac{x^2}{576}(12 - x)\left(e^{\frac{x}{6}} - 1\right)\mathrm{d}x = \frac{9e^2 - 51}{2} \approx 7.75 \text{ hours}$$

Median

The value that separates the probability distribution exactly in half.

That is, $\Pr(X \le m) = \frac{1}{2} = \Pr(X \ge m)$

$\int_{-\infty}^{m} f(x)\mathrm{d}x = \frac{1}{2}, \quad \int_{m}^{\infty} f(x)\mathrm{d}x = \frac{1}{2},$ where m is the median

Example VCAA 2014 Exam 1 Question 8a

A continuous random variable, X, has a probability density function given by x. The median of X is m.

$$f(x) = \begin{cases} \frac{1}{5}e^{-\frac{x}{5}} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

$$\int_{0}^{m} \frac{1}{5}e^{=\frac{x}{5}}\mathrm{d}x = \left[-e^{-\frac{x}{5}}\right]_{0}^{m} = -e^{-\frac{m}{5}} - (-e^{0}) = -e^{-\frac{m}{5}} + 1 = \frac{1}{2}$$

$$\Rightarrow e^{-\frac{m}{5}} = \frac{1}{2}$$

$$\Rightarrow -\frac{m}{5} = \log_e\left(\frac{1}{2}\right)$$

$$\Rightarrow m = 5\log_e(2)$$

Variance

The variance is the weighted average of the squared deviations from the mean.

$$\mathrm{var}(X) = \sigma^2 = \mathrm{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\mathrm{d}x$$

$$= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x)\mathrm{d}x = \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x + \int_{-\infty}^{\infty} -2\mu x f(x)\mathrm{d}x + \int_{-\infty}^{\infty} \mu^2 f(x)\mathrm{d}x$$

$$= \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x - 2\mu \int_{-\infty}^{\infty} -x f(x)\mathrm{d}x + \mu^2 \int_{-\infty}^{\infty} f(x)\mathrm{d}x = \mathrm{E}(X^2) - 2\mu\mathrm{E}(X) + \mu^2$$

$$= \mathrm{E}(X^2) - 2\mu^2 + \mu^2 = \mathrm{E}(X^2) - \mu^2 = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x - \mu^2$$

Variance of Linear Combinations of Random Variables

$\mathrm{var}(aX+b) = a^2\mathrm{var}(X)$, $\quad \mathrm{var}(aX+bY) = a^2\mathrm{var}(X) + b^2\mathrm{var}(Y)$

Standard Deviation

The square root of the variance. $\quad \mathrm{sd}(X) = \sigma = \sqrt{\mathrm{var}(X)}$

Standard Deviation of Linear Combinations of Random Variables

$\mathrm{sd}(aX+b) = a\,\mathrm{sd}(X)$

$\mathrm{sd}(aX+bY) = \sqrt{\mathrm{var}(aX+bY)} = \sqrt{a^2\mathrm{var}(X) + b^2\mathrm{var}(Y)}$

Example VCAA 2017 NHT Exam 2 Question 3c

The use-time of Grade B whiteboard pens is described by the probability density function f where x is the use-time in hours $f(x) = \begin{cases} \frac{x}{576}(12-x)\left(e^{\frac{x}{6}} - 1\right) & 0 \le x \le 12 \\ 0 & \text{otherwise} \end{cases}$

The expected use-time of a Grade B whiteboard pen is 7.75 hours

$$\mathrm{var}(X) = \int_0^{12} (x-7.75)^2 \frac{x}{576}(12-x)\left(e^{\frac{x}{6}} - 1\right)\mathrm{d}x \approx 5.3168$$

The standard deviation of the use-time of a Grade B whiteboard pen is $\sqrt{5.3168} \approx 2.31$ hours

Interpreting Statistics

If the standard deviation is large, the data is spread out over a widely from the mean.

If the standard deviation is small, the data is clustered close to the mean.

$\Pr(\mu - 2\sigma \le X \le \mu + 2\sigma) \approx 0.95$

Chebyshev's Theorem

At least $1 - \frac{1}{k^2}$ of the distribution lies within k standard deviations of the mean.

$1 - \frac{1}{k^2} \le \Pr(\mu - k\sigma \le X \le \mu + k\sigma) \le 1, \qquad k > 1$

For $k = 2$, $0.75 \le \Pr(\mu - 2\sigma \le X \le \mu + 2\sigma) \le 1$