Random variables

Table of contents

In the real world, variables change due to various factors - thus have some *random* component. Thus, we treat the measurements (variables) as *random variables*.

> The random variable X can be defined as a *function* which assigns a *number* to each outcome in the **sample space**.

A random variable is characterised by it's probabilities, and how they vary over different values of x. Thus, for a discrete experiment, of a random variable X defined in S we get:

$$\mathbb{P}(X = x) = \sum_{s:X=x} P(\{s\})$$

Example: Imagine we toss a fair coin 3 times. Let X denote the number of heads that turned up. Then

$$f_X(0) = \mathbb{P}(X = 0) = \mathbb{P}(\{\text{TTT}\}) = \frac{1}{8}$$
$$\ldots$$

Discrete Random Variables

> The random variable X is **discrete** if there are countably many values x for which $\mathbb{P}(X = x) > 0$.

Probability function

> The **probability function** of the discrete random variable X is the function f_X given by:
>
> $$f_X(x) \equiv \mathbb{P}(X = x)$$
>
> where
>
> $$x \in \mathbb{R} \text{ and } \mathbb{P}(X = x) > 0$$

Thus, it follows that $\sum f_X(x) = 1$, and that $f_X(x) >= 0$.

Expected Value

The expected value, or *mean* for a discrete random variable X is defined by:

> $$\mathbb{E}(X) = \sum_{\text{all } x} x \cdot \mathbb{P}(X = x) = \sum_{\text{all } x} x \cdot f_X(x)$$

Example: Let X be the number of females in a commitee with three members.
Assume that there is a 50:50 chance of each committee member being a female, and that committee members are chosen independently of eachother.

Find $\mathbb{E}(X)$.

Therefore:

$$f_X(0) = 1/8, f_X(1) = 3/8, f_X(2) = 3/8, f_X(3) = 1/8$$

Thus, the total answer is:

$$0 \cdot 1/8 + 1 \cdot 3/8 + 2 \cdot 3/8 + 3 \cdot 1/8 = 1.5$$

Continuous random variables

Density functions (PDF)

The analogue of the probability function for continuous random variables is the **density function** or **probability density function**.

> The **density function** of a continuous random variable is a real-valued function f_X on $\mathbb{R}$ with the property
>
> $$\int_A f_X(x)dx = \mathbb{P}(X \in A)$$
>
> for any (measurable set) $A \subseteq R$.

Thus, the following properties follow:

- $f_X(x) >= 0$ for all $x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$

It also follows that the probability over some pair of numbers a and b within the random variable X is given by:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx$$

Example: The lifetime (in thousands of hours) X of a light bulb has density $f_X(x) = e^{-x}, x > 0$. Find the probability that a light bulb lasting between 2 thousand and 3 thousand hours.

$$\begin{aligned}\int_2^3 e^{-x}dx &= [-e^{-x}]_2^3 \\ &= -e^{-3} + e^{-2} \\ &= 0.0855\ldots\end{aligned}$$

For continuous random variables, we have the property:

$$\mathbb{P}(X = a) = 0$$

for any $a \in \mathbb{R}$. For a single point, there is no probability/no logical understanding of probability. Rather, we look at some subset of X to gain insights. Thus, it follows that:

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \ldots$$

Cumulative distribution function (CDF)

> The **cumulative distribution function** (cdf) of the random variable X is
>
> $$F_x(x) = \mathbb{P}(X \leq x)$$

This definition applies to both discrete and continuous variables. Thus, we are seeing the accumulation of properties as we increase in the random variable X. From the above definition, it follows that:

$$F_x(b) - F_x(a) = \mathbb{P}(a < X \leq b)$$

The CDF can be formally defined as:

> $$ F*X(x) = \begin{cases} \sum*{t \le x} f*X(t) & \text{if X is discrete} \ \int*{-\infty}^{x}f_X(t)dt & \text{if X is continuous} \end{cases}

$Furthermore :> \$\$ f_X(x) = \begin{cases} F_X(x) - F_X(x_-) & \text{if } X \text{ is discrete} \\ F'_X(x) & \text{if } X \text{ is continuous} \end{cases} \$\$ **Example** : Acoinwith\$p\$ =$

Example: A point is chosen at random inside a circle that has radius r. Let X be the distance from the centre of the circle. Find $f_X(x)$, the density function of X, where the point sits within x.

$$\begin{aligned} f_X(x) &= \frac{2\pi x}{\pi r^2} \\ &= \frac{2x}{r^2} \end{aligned}$$

Quantiles

If F_X is strictly increasing in some interval, then F_X^{-1} is well defined, and for a specified $p \in (0, 1)$, the pth quantile of F_X is x_p where

$$F_X(x_p) = p$$

or

$$x_p = F_X^{-1}(p)$$

Example: Let X be a random variable with cumulative distribution function

$$F_X(x) = 1 - e^{-x}, x > 0$$

find the median and quartiles of X.

Since $F_X(x) = 1 - e^{-x}$, then $y = F_X(x)$, thus $y = 1 - e^{-x}$; thus for $x = F_x^{-1}(y) = -\ln(1-y)$, thus

$$x_p = -\ln(1-p)$$

Thus, median = $-\ln(2)$, etc.

Expected Value

The expected value, or mean of a continuous random variable X, is given by:

> $$\mathbb{E}(X = x) = \int_{-\infty}^{\infty} x f_X(x) dx$$
>
> where $f_X(x)$ is the density function.

Thus, $\mathbb{E}(X)$, can be considered the *fulcrum* or *centre of gravity* of your data/histogram - as it is the **long run average**.

Effects of a linear transformation

$$\begin{aligned}
\mathbb{E}(a + bX) &= \sum_x (a + bx)\mathbb{P}(X = x) \\
&= a\sum_x \mathbb{P}(X = x) + b\sum_x x\mathbb{P}(X = x) \\
&= a + b\mathbb{E}(X)
\end{aligned}$$

Example: Suppose a random variable X has a density function

$$f_X(x) = e^{-x}, x > 0$$

Find $\mathbb{E}(X)$

$$\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} x \cdot e^{-x} \\
&= \left[xe^{-x}\right]_0^{\infty} + \int_0^{\infty} e^{-x} dx \\
&= [0 - 0] + \left[e^{-x}\right]_0^{\infty} \\
&= 1
\end{aligned}$$

Transformation of random variables

If some function g transforms our random variable X, then $g(X)$ is a transformed random variable. The expected value of the function $g(X)$ can be given by:

$$\mathbb{E}(g(X)) = \begin{cases} \sum_{\text{all } x} g(x) f_X(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x) f_X(x) & \text{if } X \text{ is continuous} \end{cases}$$

For a sum:

$$\mathbb{E}\{g_1(X) + g_2(X) + \ldots\} = \mathbb{E}\{g_1(X)\} + \mathbb{E}\{g_2(X)\} + \ldots$$

Moments

> The r-th moment about some constant a, is defined by
>
> $$\mathbb{E}(X-a)^r$$
>
> The first moment is mean, the second moment is variance, and so on. Moments are very useful.

Variance and Standard Deviation

Thus, the variance is now defined the be:

$$\mathrm{Var}(X) \equiv \sigma_X^2 = \mathbb{E}\{(X-\mu)^2\}$$

Thus, the variance is the second moment of X about μ. Variances are generally easier to interpret mathematically, and are more applied in other notations - thus we use them more; but standard deviation is a better physical indicator of spread.

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2$$

is another definition of variance (and the one generally used in high school). Note that $\sigma = \sqrt{Var(X)}$, as defined in previous chapters.

$$\mathrm{Var}(a+bX) = b^2 \mathrm{Var}(X)$$
$$\sigma(a+bX) = |b| \sigma(X)$$

Independent random variables

In many cases, we have more than one variable we have to consider. For future examples, let us consider two random variables, denoted by X and Y - in this case, we have the case of *bivariate random variables*.

Newly defined rules for bivariate random variables

Since bivariate variables are independent; we can model them in linear transformations. If the variables X and Y were not independent, then it would be useful to consider the *co-variance*, which is a measure of parity in movement.

Note: the $\pm$ must follow for the entire equation

$$\mathbb{E}(aX \pm bY) = a\mathbb{E}(X) \pm b\mathbb{E}(Y)$$

$$\mathrm{Var}(aX \pm bY + c) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y)$$

Useful identities

Chebychev's Inequality

Chebychev's ineqlauty is a fundamental result concerning **tail probabilities** of general random variables. It is useful for derivation of convergence results given later in the notes.

> **Chebychev's Inequality:** If X is any random variable with $\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$ then
>
> $$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$
>
> where $k > 0$ is a constant.
>
> More simply, we can describe it as: *the probability (range) that X is more than k standard deviations from it's mean*.

Proof:

$$\begin{aligned}\sigma^2 = \mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx \\ &\geq \int_{|x-\mu|>k\sigma} (x-\mu)^2 f_X(x)dx \\ &\geq \int_{|x-\mu|>k\sigma} (k\sigma)^2 f_X(x)dx\end{aligned}$$

Since $|x-\mu| > k\sigma \implies (x-\mu)^2 f_X(x) > (k\sigma)^2 f_X(x)$

$$\begin{aligned}\sigma^2 &\geq k^2\sigma^2 \int_{|x-\mu|>k\sigma} f_X(x)dx \\ &= k^2\sigma^2\mathbb{P}(|X-\mu| > k\sigma)\end{aligned}$$

$$\therefore \mathbb{P}(|X-\mu| > k\sigma) \leq \frac{1}{k^2}$$

Example: A factory produces items per day with mean 500 and variance 100. Whats the lower bound for the probability that between 400 and 600 items will be produced tomorrow?

Therefore, we are finding a such that $P(400 \leq X \leq 600) \geq a$. Note that $\mu = 500$ and $\sigma^2 = 100, \sigma = 10$. Therefore, we have $P(400 \leq X \leq 600) = 1 - \{P(X < 400 + P(X > 600)\}$.

$$
\begin{aligned}
&= 1 - \{P(X - \mu < 400 - \mu) + P(X - \mu > 600 - \mu)\} \\
&= 1 - \{P(X - \mu < -100) + P(X - \mu > 100)\} \\
&= 1 - P(|X - \mu| > 100) \\
&= 1 - P(|X - \mu| > k\sigma), k = 10 \\
&\implies 1 - \frac{1}{10^2} = 1 - \frac{1}{100} = \frac{99}{100}
\end{aligned}
$$

Markov Inequality

> **Markov inequality:** For any non-negative random variable X and $a > 0$
>
> $$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Jensen's Inequality

> **Jensen's Inequality:** Suppose h is a convex functionn and X is a random variable then:
>
> $$h(\mathbb{E}(X)) \leq \mathbb{E}(h(X))$$
>
> Suppose h is a concave function and X is a random variable, then:
>
> $$h(\mathbb{E}(X)) \geq \mathbb{E}(h(X))$$

Transformation of random variables

We are generally more interested how the variable becomes distributed after a transformation. For a random variable X, with known $f_X(x)$, we are now interested in the distibribution of $Y = h(X)$. Thus, we are interested in finding $f_Y(y)$.

Discrete

> Let $Y = h(X)$, then for discrete X, we have:
>
> $$f_Y(y) = P(Y = y) = P\{h(X) = y\} = \sum_{x:h(x)=y} P(X = x)$$

Continuous

> Let $Y = h(X)$, then for a continus random variable X, if h is monotonic over the set $\{x : f_X(x) > 0\}$, then:

> $$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$
> $$= f_X\{h^{-1}(y)\}\left|\frac{dx}{dy}\right|$$
>
> We apply $h^{-1}(y)$ (which can be found by solving for X), to have it in terms of Y.

Example: Let

$$f_X(x) = 3x^2, 0 < x < 1$$

Find $f_Y(y)$ where $Y = 2X - 1$.

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

$$= 3\left(\frac{y+1}{2}\right)^2\frac{1}{2}$$

$$\frac{3}{2}\left(\frac{y+1}{2}\right)^2$$