Math Refresher: Basic Statistics and Probability

Random Variables

A random variable is a variable whose value is determined by the outcome of a random experiment. For example, if we toss a coin, the outcome is random, but the possible values of X are 0 and 1. If we roll a die, the outcome is random with possible values 1, 2, 3, 4, 5, and 6.

There are two kinds of random variables:

- **Discrete random variables** can only take on a finite number of values. For example, the number of heads in 10 coin tosses is a discrete random variable.
- **Continuous random variables** can take on any value in a range. For example, the height of a randomly selected person is a continuous random variable.

If X is discrete random variable, then $P(X = c)$ is the probability that X takes on the value c. It can be any value between 0 and 1.

By definition, the sum of all probabilities for all feasible values of X is 1. That is, $\sum_c P(X = c) = 1$.

If X is continuous random variable, then $P(X = c) = 0$ for any value c. The probability to observe a particular number is zero. Instead, when using continuous data, we focus on the probability of observing a value in a range. For example, $P(1.7 \leq X \leq 1.8)$ is the probability that X is between 1.7 and 1.8, which can be any value between 0 and 1.

Probability Distributions

A probability distribution is a function that assigns probabilities to the values of a random variable. For discrete random variables, we can use a table to describe the probability distribution. For example, the probability distribution of the number of heads in 5 coin tosses is:

Number of heads	Probability
0	0.03125
1	0.15625
2	0.3125
3	0.3125
4	0.15625
5	0.03125

In this case, the sum of all probabilities is 1.

For continuous random variables, we can use a function to describe the probability distribution. For example, we can say that the probability distribution of the height of a randomly selected person is:

$$f(x)$$

This function has important properties:

- $f(x) \geq 0$ for all x.
- $\int_{-\infty}^{\infty} f(x)dx = 1$.
- $P(a \leq X \leq b) = \int_a^b f(x)dx$.
- $P(X \leq a) + P(X > a) = 1$.
- $P(a \leq X \leq b) = P(X < b) - P(X < a)$.

Joint Probability Distributions

The joint probability distribution of X and Y is a function that assigns probabilities to the values of X and Y. For discrete random variables, we can use a table to describe the joint probability distribution. For continuous variables, it must be the case that:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$$

Marginal Probability Distributions

The marginal probability distribution of X is the probability distribution of X ignoring the values of Y. This can be expressed as:

$$P(x) = \sum_{z=-\infty}^{\infty} P(x, y = z)$$

it still must be the case that

$$\sum_{w=-\infty}^{\infty} \sum_{z=-\infty}^{\infty} P(x = w, y = z) = 1$$

For continuous random variables, we have the following

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

where $f(x)$ is the marginal probability distribution of X. What is left after we "integrate out" Y is the marginal probability distribution of X.

$$f(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

Independence

Two random variables X and Y are independent if and only if:

$$P(x, y) = P(x)P(y) or f(x, y) = f(x) * f(y)$$

Conditional Probability

The conditional probability of X given Y is:

$$P(x|y) = \frac{P(x, y)}{P(y)}$$

or, the conditional probabilty density function:

$$f(x|y) = \frac{f(x, y)}{f(y)}$$

And if X and Y are independent, then:

$P(x|y) = P(x)$ or $f(x|y) = f(x)$.

Mean, and variance

The mean of a random variable X is:

$$E(X) = \sum_x xP(x)$$

or

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

Which is a weighted sum of all possible values of X, and where the weights are the probabilities (or densities) of each value. It can also be written or referred as:

$$E(X), \mu_x, \bar{x}$$

This measure is also called the **expected value** of X, and proviveds a measure of the "center" of the distribution of X. It can be very sensitive to outliers.

The variance of a random variable X is:

$$Var(X) = E[(X - E(X))^2]$$

$$Var(X) = \sum_x (X - E(X))^2 P(x)$$

or

$$Var(X) = \int_x (X - E(X))^2 f(x)dx$$

Which is the expected value of the squared difference between X and its mean. It provides a measure of average the "spread" of the distribution of X.

It could also be defined as follows:

$$\sigma_x^2 = Var(x) = E(X^2) - [E(X)]^2$$

There are other measures that can be used to characterize a distribution, such as the median, the mode, the skewness, and the kurtosis. They are defined as follows:

- The **median** is the value of X such that $P(X \leq x) = 0.5$.
- The **mode** is the value of X that maximizes $P(X = x)$.
- The **skewness** is a measure of the asymmetry of the distribution of X. It is defined as:

$$\frac{E[(X-E(X))^3]}{[Var(X)]^{3/2}}$$

- The **kurtosis** is a measure of the "peakedness" of the distribution of X. It is defined as:

$$\frac{E[(X-E(X))^4]}{[Var(X)]^2}$$

- The **quantiles** of a distribution are values that divide the distribution into equal parts. For example, the 0.25 quantile is the value of X such that $P(X \leq x) = 0.25$.

For a **normal distribution**, the mean, median, and mode are all equal. The skewness is 0, and the kurtosis is 3.

Covariance and Correlation

The covariance of two random variables X and Y is:

$$Cov(X,Y) = E[(X-E(X))(Y-E(Y))]$$

$$Cov(X,Y) = \sum_x \sum_y (x-E(X))(y-E(Y))P(x,y)$$

$$Cov(X,Y) = \int_x \int_y (x-E(X))(y-E(Y))f(x,y)$$

The covariance measures the **linear** association between X and Y. If X and Y are independent, then $Cov(X,Y) = 0$. However, if $Cov(X,Y) = 0$, then X and Y are not necessarily independent. For example $y = (x-E(X))^2$ and x are not independent, but $Cov(y,x) = 0$.

This measure is scale dependent. For example, if we measure X in meters, and Y in centimeters, then $Cov(X,Y)$ will be 100 times larger than if we measure X in meters and Y in kilometers.

An alternative measure of association is the correlation coefficient, which is defined as:

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_x \sigma_y}$$

This statistics is always between -1 and 1, regardless of the scale of x or y.

Propeties of Mean, Variance and Covariance

Consider two random variables X and Y, and let a, b, c and d be constants. Then:

- $Var(aX + b) = a^2 Var(X)$
- $Cov(aX + b, cY + d) = acCov(X, Y)$
- $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$
- $Cov(X, Y) = E(XY) - E(X)E(Y)$
- $Cov(X, X) = Var(X)$

For the mean:

- $E(aX + b) = aE(X) + b$
- $E(aX + bY) = aE(X) + bE(Y)$

Some useful distributions

Discrete distributions

- **Bernoulli distribution**: $X \sim Bernoulli(p)$, where $p = P(X = 1)$ and $1-p = P(X = 0)$. $E(X) = p$ and variance $Var(X) = p(1 - p)$. Flip a coin with probability p of getting heads.
- **Binomial distribution**: $X \sim Binomial(n, p)$, where $p = P(X = 1)$ and $1-p = P(X = 0)$. $E(x) = np$ and $Var(X) = np(1 - p)$. The binomial distribution is the distribution of the number of successes in n independent Bernoulli trials.
- **Poisson distribution**: $X \sim Poisson(\lambda)$, where $\lambda = E(X) = Var(x)$. Typically used for counts. For example, the number of customers arriving at a store in a given hour.

Continuous distributions

- **Uniform distribution**: $X \sim Uniform(a, b)$, where $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$, and $f(x) = 0$ otherwise. $E(X) = \frac{a+b}{2}$ and $Var(X) = \frac{(b-a)^2}{12}$. Time between bus arrivals.
- **Normal distribution**: $X \sim Normal(\mu, \sigma^2)$, where $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. $E(X) = \mu$ and $Var(X) = \sigma^2$. For example, the height of a randomly selected person.
- **t-distribution**: $X \sim t(\nu)$, where $f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}(1 + \frac{x^2}{\nu})^{-\frac{\nu+1}{2}}$. $E(X) = 0$ if $\nu > 1$, and $Var(X) = \frac{\nu}{\nu-2}$ if $\nu > 2$. For example, the distribution of the sample mean of a small sample from a normal distribution.

Alternatively. $X \sim t(\nu)$, where $X = \frac{Z}{\sqrt{V/\nu}}$, where $Z \sim Normal(0, 1)$ and $V \sim \chi^2(\nu)$, and Z and V are independent.

- **Chi-squared distribution**: $X \sim \chi^2(\nu)$, where $f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}$. $E(X) = \nu$ and $Var(X) = 2\nu$.

  Alternatively, $X \sim \chi^2(\nu)$, where $X = Z_1^2 + Z_2^2 + ... + Z_\nu^2$, where $Z_i \sim Normal(0, 1)$, and Z_1, Z_2, ..., Z_ν are independent.

- **F-distribution**:

  $X \sim F(\nu_1, \nu_2)$, where $f(x) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} (\frac{\nu_1}{\nu_2})^{\nu_1/2} x^{\nu_1/2-1} (1 + \frac{\nu_1}{\nu_2} x)^{-(\nu_1+\nu_2)/2}$. $E(X) = \frac{\nu_2}{\nu_2 - 2}$ if $\nu_2 > 2$, and $Var(X) = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$ if $\nu_2 > 4$.

  Alternatively, $X \sim F(\nu_1, \nu_2)$, where $X = \frac{V_1/\nu_1}{V_2/\nu_2}$, where $V_1 \sim \chi^2(\nu_1)$ and $V_2 \sim \chi^2(\nu_2)$, and V_1 and V_2 are independent.