Mathematical Expectation.

Definition: Expectation X of having values $x_1, \dots, x_n$ is:

$$E(X) = x_1 P(X=x) + \dots + x_n P(X=x) = \sum_{j=1}^{n} x_j P(X=x_j), \text{ or}$$

if $P(X=x_j) = f(x_j)$:

$$E(X) = x_1 f(x_1) + \dots + x_n f(x_n) = \sum_{j=1}^{n} x_j f(x_j) = \sum x f(x).$$

If all probabilities equal:

$$E(X) = (x_1 + \dots + x_n)/n.$$

If X takes infinite number of values $x_1, x_2, \dots$

$E(X) = \sum_{j=1}^{\infty} x_j f(x_j)$ provided the infinite series converges.

Continuous case: $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$ provided integral converges absolutely.

Example: Dice rules: \$20 for 2, \$40 for 4, \$30 loss for 6, and neither wins nor loses for anything else.

$$E(X) = (0)\left(\tfrac{1}{6}\right) + (20)\left(\tfrac{1}{6}\right) + (0)\left(\tfrac{1}{6}\right) + (40)\left(\tfrac{1}{6}\right) + (0)\left(\tfrac{1}{6}\right) + (-30)\left(\tfrac{1}{6}\right) = 5.$$

e.g. player can expect to win \$5.

x_j	0	+20	0	+40	0	−30
$f(x_j)$	1/6	1/6	1/6	1/6	1/6	1/6

Example: Density function of random variable X is $f(x) = \begin{cases} x/2 & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$

The expected value of X is $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$

$$= \int_0^2 x\left(\tfrac{1}{2}x\right)dx = \int_0^2 \frac{x^2}{2}\,dx = \left.\frac{x^3}{6}\right|_0^2 = 4/3.$$

Functions: X discrete random variable with probability $f(x)$, then $Y = g(X)$ is also discrete random variable, probability function of Y is:

$$h(y) = P(Y=y) = \sum_{\{x \mid g(x)=y\}} P(X=x) = \sum_{\{x \mid g(x)=y\}} f(x)$$

If X takes on values $x_1, x_2, \ldots, x_n$ and Y the values $y_1, y_2, \ldots, y_m$ $(m \geq n)$ then $y_1 h(y_1) + y_2 h(y_2) + \ldots + y_m h(y_m) = g(x_1)f(x_1) + \ldots + g(x_n)f(x_n)$

$$\therefore E[g(X)] = g(x_1)f(x_1) + \ldots + g(x_n)f(x_n) = \sum_{j=1}^{n} g(x_j)f(x_j) = \sum g(x)f(x).$$

Similarly in the continuous case:

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f(x,y)\, dx\, dy.$$

Example:

$$E(3X^2 - 2X) = \int_{-\infty}^{\infty} (3x^2 - 2x) f(x)\, dx = \int_0^2 (3x^2 - 2x)\left(\tfrac{1}{2}x\right) dx = 10/3.$$

where $f(x)$ is the previous example

Theorems:

1. If c any constant, $E(cX) = cE(X)$.
2. If X, Y any random variables then $E(X+Y) = E(X) + E(Y)$.
3. $E(XY) = E(X)E(Y)$

Variance and Standard Deviation:

Mean of X is μ.

Variance is $Var(X) = E[(X-\mu)^2]$, the positive square root is standard deviation.

$$\sigma_X = \sqrt{Var(X)} = \sqrt{E[(X-\mu)^2]}$$

If X discrete random variable with values $x_1, x_2, \ldots, x_n$, the variance is:

$$\sigma_X^2 = E[(X-\mu)^2] = \sum_{j=1}^{n} (x_j - \mu)^2 f(x_j) = \sum (x-\mu)^2 f(x).$$

In special case where probabilities equal:

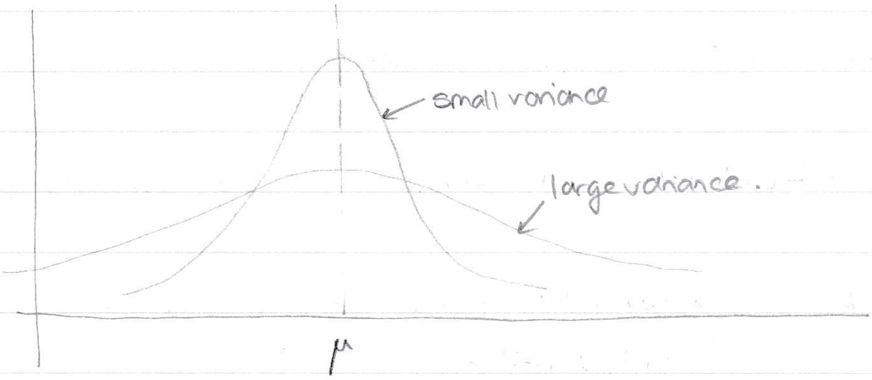
$$\sigma^2 = [(x_1-\mu)^2 + \ldots + (x_n-\mu)^2] / n$$

If X takes infinite number of values $x_1, x_2, \dots$ and series converges:

$$\sigma_X^2 = \sum_{j=1}^{\infty} (x_j - \mu)^2 f(x_j).$$

Similarly, for continuous case:

$$\sigma_X^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$; again provided convergence.



Example: Variance and standard deviation for previous example

We found mean $\mu = E(X) = 4/3$, so variance is:

$$\sigma^2 = E\left[\left(X - \tfrac{4}{3}\right)^2\right] = \int_{-\infty}^{\infty} \left(x - \tfrac{4}{3}\right)^2 f(x)\,dx$$

$$= \int_0^2 \left(x - \tfrac{4}{3}\right)^2 \left(\tfrac{1}{2}x\right) dx = 2/9.$$

standard deviation is $\sqrt{2/9} = \sqrt{2}/3$.

Theorems:

1. $\sigma^2 = E[(X-\mu)^2] = E(X^2) - [E(X)]^2$, $\mu = E(X)$.

2. If c constant, $Var(cX) = c^2 Var(X)$

3. The quantity $E[(X-a)^2]$ is a minimum when $a = \mu = E(X)$

4. If X and Y are independent random variables:
   $Var(X+Y) = Var(X) + Var(Y)$ or $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$
   $Var(X-Y) = Var(X) + Var(Y)$ or $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$.

Standardised Random Variables: Let X be random variable with mean and standard deviation $(\sigma > 0)$, the **standardized random variable** is given by:

$$X^* = \frac{X-\mu}{\sigma}, \quad \text{which has a mean and variance of } 1.$$

i.e. $E(X^*) = 0$ and $Var(X^*) = 1$.

Moments The r^{th} moment of a random variable X about the mean is

$$\mu_r = E[(X-\mu)^r], \quad r = 0,1,2,\dots \quad \text{and } \mu_0 = 1, \mu_1 = 0, \mu^2 = \sigma^2$$

Discrete: $\mu_r = \sum (x-\mu)^r f(x)$

Continuous: $\mu_r = \int_{-\infty}^{\infty} (x-\mu)^r f(x)\,dx$.

r^{th} moment about origin:

$$\mu_r' = E(X^r).$$

$$\mu_r = \mu_r' - \binom{r}{1}\mu_{r-1}'\mu + \dots + (-1)^j\binom{r}{j}\mu_{r-j}'\mu^j + \dots + (-1)^r\mu_0'\mu^r$$

As special cases, $\mu_1' = \mu$ and $\mu_0' = 1$

$$\mu_2 = \mu_2' - \mu_2$$
$$\mu_3 = \mu_3' - 3\mu_2'\mu + 2\mu^3$$
$$\mu_4 = \mu_4' - 4\mu_3'\mu + 6\mu_2'\mu^2 - 3\mu^4$$

Moment Generating Function: $M_X(t) = E(e^{tX})$, and assuming convergence:

$$M_X(t) = \sum e^{tx} f(x) \quad \text{(discrete)}$$
$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx \quad \text{(continuous)}.$$

From a **Taylor Series** expansion:

$$M_X(t) = 1 + \mu t + \mu_2' \frac{t^2}{2!} + \dots + \mu_r' \frac{t^r}{r!} + \dots$$

Also, $\mu_r' = \frac{d^r}{dt^r} M_X(t) \Big|_{t=0}$, where μ_r' is the r-th derivative of $M_X(t)$ evaluated at $t=0$

Theorems:

1. If $M_X(t)$ moment generating function of random variable X and a and b $(b \neq 0)$ then
$$M_{(X+a)/b}(t) = e^{at/b} M_X(t/b).$$
2. If X and Y independent random variables having moment generating functions $M_X(t)$ and $M_Y(t)$
$$M_{X+Y}(t) = M_X(t) M_Y(t).$$
3. Suppose X and Y are random variables with $M_X(t)$ and $M_Y(t)$.
Then X and Y have the same probability distribution iff $M_X(t) = M_Y(t)$.

Characteristic Function:

Let $t = i\omega$, the characteristic function is:

$$\phi_X(\omega) = M_X(i\omega) = E(e^{i\omega X})$$

It follows:

$$\phi_X(\omega) = \sum e^{i\omega x} f(x), \quad \text{discrete}$$
$$\phi_X(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} f(x)\,dx, \quad \text{continuous}$$
(as $|e^{i\omega x}| = 1$, the series / integral converges)

$$\phi_X(\omega) = 1 + i\mu\omega - \mu_2' \frac{\omega^2}{2!} + \dots + i^r \mu_r' \frac{\omega^r}{r!} + \dots$$ where

$$\mu_r' = (-1)^r i^r \frac{d^r}{d\omega^r} \phi_X(\omega) \Big|_{\omega=0}.$$

Theorem: If $\phi_X(\omega)$ characteristic function of random variable X and a and b $(b \neq 0)$ are constants, then characteristic function of $(X+a)/b$ is:

$$\phi_{(X+a)/b}(\omega) = e^{ai\omega/b} \phi_X(\omega/b).$$

Theorem: If X and Y are **independent** random variables with characteristic functions $\phi_X(\omega)$ and $\phi_Y(\omega)$,

$$\phi_{X+Y}(\omega) = \phi_X(\omega)\,\phi_Y(\omega)$$

Theorem: Suppose X, Y random variables having characteristic functions $\phi_X(\omega)$ and $\phi_Y(\omega)$.
X and Y have the **same probability distribution** iff $\phi_X(\omega) = \phi_Y(\omega)$.

Inversion Formula:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega x}\,\phi_X(\omega)\,d\omega.$$

Variance for Joint Distributions: If X, Y are two **continuous** random variables having **joint** density function $f(x,y)$, the means, or **expectations** of X and Y are:

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

$$\mu_Y = E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y)\,dx\,dy.$$

The **variances** are:

$$\sigma_X^2 = E[(X-\mu_X)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)^2 f(x,y)\,dx\,dy$$

$$\sigma_Y^2 = E[(Y-\mu_Y)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y-\mu_Y)^2 f(x,y)\,dx\,dy.$$

Covariance:

$$\sigma_{XY} = Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

In terms of **joint density** function:

$$\sigma_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f(x,y)\,dx\,dy.$$

Similarly for two discrete random variables

$$\mu_X = \sum_x \sum_y x f(x,y) \qquad \mu_Y = \sum_x \sum_y y f(x,y) \qquad \sigma_{XY} = \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f(x,y).$$

Theorem: $\sigma_{XY} = E(XY) - E(X)E(Y) = E(XY) - \mu_X \mu_Y$

Theorem: If X and Y are independent random variables then $\sigma_{XY} = Cov(X,Y) = 0$

Theorem: $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X,Y)$.

$$\sigma^2_{X \pm Y} = \sigma_X^2 + \sigma_Y^2 \pm 2\sigma_{XY}.$$

Theorem: $|\sigma_{XY}| \leq \sigma_X \sigma_Y$.

Correlation Coefficient: If X and Y are independent then $Cov(X,Y) = \sigma_{XY} = 0$

Otherwise, if X and Y are completely dependent, i.e. when $X = Y$, then $Cov(X,Y) = \sigma_{XY} = \sigma_X \sigma_Y$.

The measure of dependence of X and Y is:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad \text{(correlation coefficient).} \qquad -1 \leq \rho \leq 1.$$

when $\rho = 0$, X and Y are unrelated

Conditional Expectation, Variance, and Moments: If X, Y have joint density function $f(x,y)$, the conditional density function of Y given X is $f(y|x) = f(x,y)/f_1(x)$ where $f_1(x)$ is the marginal density function of X.

Conditional expectation, or conditional mean, of Y given X by

$$E(Y|X=x) = \int_{-\infty}^{\infty} y f(y|x)\,dy.$$

where "$X = x$" is interpreted as $x < X \leq x + dx$ in the continuous case.

Properties:

1. $E(Y|X=x) = E(Y)$ when X and Y are independent.

2. $E(Y) = \int_{-\infty}^{\infty} E(Y|X=x) f_1(x)\,dx$

Chebyshev's ~~Example~~: Inequality:

Suppose X is a random variable with **mean** μ and **variance** σ^2, which are finite.
Then ϵ is any **positive number**:

$$P(|X-\mu| \geq \epsilon) \leq \sigma^2/\epsilon^2, \quad \text{or with } \epsilon = k\sigma,$$
$$P(|X-\mu| \geq k\sigma) \leq 1/k^2.$$

Example:

Letting $k=2$:

$$P(|X-\mu| \geq 2\sigma) \leq 0.25 \quad \text{or} \quad P(|X-\mu| < 2\sigma) \geq 0.75.$$

i.e. the **probability** of X differing from its mean by more than **2 standard deviations** is **less** than or equal to **0.25**.
Equivalently, the probability that X will lie **within** 2 standard deviations of its mean is greater than or equal to 0.75.

Law of Large Numbers:

Let $X_1, X_2, \ldots, X_n$ be **mutually independent** random variables, each having finite **mean** μ and **variance** σ_2.
Then if $S_n = X_1 + X_2 + \ldots + X_n$ $(n = 1, 2, \ldots)$:

$$\lim_{n \to \infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0.$$

Mode:

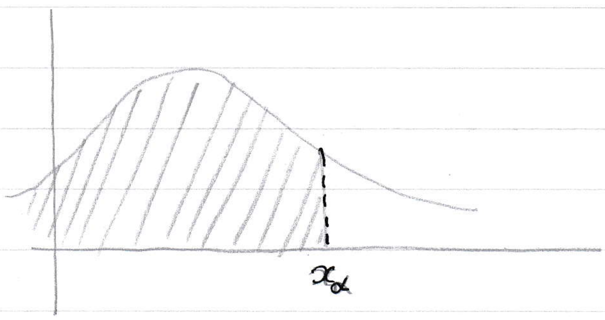
Of a discrete random variable is the **value** which occurs **most often**.

Median:

The **median** is that value x for which $P(X<x) \leq \frac{1}{2}$ and $P(X>x) \leq \frac{1}{2}$.
In case of **continuous**, $P(X<x) = \frac{1}{2} = P(X>x)$

Percentiles:

Subdivide the area under a density curve, i.e. $x_{0.10}$ is 10%.



Semi Interquartile Range: If $x_{0.25}$ and $x_{0.75}$ are the 25th and 75th **percentile values**, the **difference** $x_{0.75} - x_{0.25}$ is the **interquartile range**.
The **semi-interquartile range** is $\frac{1}{2}(x_{0.75} - x_{0.25})$.

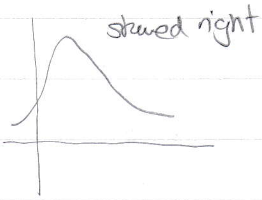
Mean Deviation: The **mean deviation** of a random variable X is the **expectation** of $|X-\mu|$, assuming **convergence**:

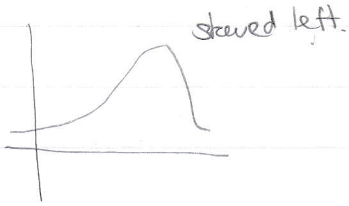
$$M.D.(X) = E[|X-\mu|] = \sum |x-\mu| f(x), \quad \text{discrete}.$$

$$M.D.(X) = E[|X-\mu|] = \int_{-\infty}^{\infty} |x-\mu| f(x)dx.$$

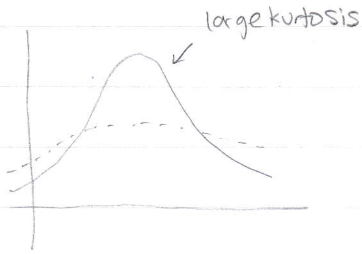
Skewness: A **non-symmetric distribution**, where "skewed to the right" means a longer tail on the left
The **coefficients** of **skewness**, one measure is given by:

$$\alpha_3 = \frac{E[(X-\mu)^3]}{\sigma^3} = \frac{\mu_3}{\sigma^3}$$





Kurtosis: Where a distribution's values are **concentrated** in the **middle**



Coefficient of **kurtosis**:

$$\alpha_4 = \frac{E[(X-\mu)^4]}{\sigma^4} = \frac{\mu_4}{\sigma^4}.$$