Expectation

* Expectation of a Random Variable

⇒ The expected value, or mean, or first moment, of X is defined to be

$$E(X) = \int x\, dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x)\, dx & \text{if } X \text{ is continuous} \end{cases}$$

assuming that the sum (or integral) is well defined.

↳ We use the following notation to denote the expected value of X.

$$\mathbb{E}(X) = \mathbb{E}X = \int x\, dF(x) = \mu = \mu_X$$

Theorem: (The Rule of the Lazy Statistician)

Let $Y = r(X)$, Then

$$\mathbb{E}(Y) = \mathbb{E}(r(X)) = \int r(x)\, dF_X(x)$$

⇒ Let A be an event and let $r(x) = I_A(x)$

where $I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$

$$\boxed{\mathbb{E}(I_A(X)) = \int I_A(x) f_X(x)\, dx = \int_A f_X(x)\, dx = \mathbb{P}(X \in A)}$$

⇒ Functions of Several variables are handled in a similar way.

↳ If $Z = g(X, Y)$

$$\mathbb{E}(Z) = \mathbb{E}(g(X,Y)) = \iint g(x,y)\, dF(x,y)$$

⇒ The K^{th} moment of X is defined to be $\mathbb{E}(X^K)$ assuming that $\mathbb{E}(|X|^K) < \infty$

↳ If the K^{th} moment exist and if $j < K$ then the j^{th} moment exists.

★ <u>Properties of Expectations</u>

<u>Theorem:</u> If $X_1, \dots X_n$ are random variables and $a_1, \dots a_n$ are Constants, then

$$\mathbb{E}\left(\sum_i a_i X_i\right) = \sum_i a_i \mathbb{E}(X_i)$$

<u>Theorem:</u> Let $X_1, \dots X_n$ be Independent random variable then,

$$\boxed{\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_i \mathbb{E}(X_i)}$$

★ Variance and Covariance

Let X be a random variable with mean μ. The <u>variance</u> of X - denoted by σ^2 or σ_x^2 or $\mathbb{V}(X)$ or $\mathbb{V}X$ is defined by

$$\sigma^2 = \mathbb{E}(X-\mu)^2 = \int (x-\mu)^2 \, dF(x)$$

assuming this expectation exists.

↳ The <u>standard deviation</u> $sd(X) = \sqrt{\mathbb{V}(X)}$ and is also denoted by σ & σ_x.

⇒ Variance has the following properties:

1. $\mathbb{V}(X) = \mathbb{E}(X^2) - \mu^2$
2. If a and b are constants then
$$\mathbb{V}(aX+b) = a^2 \mathbb{V}(X)$$
3. If $X_1, \ldots X_n$ are independent and $a_1, \ldots a_n$ are Constants, then

$$\mathbb{V}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathbb{V}(X_i)$$

⇒ If $X_1, \ldots X_n$ are random variables then we define <u>sample mean</u> to be

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and the <u>sample variance</u> to be

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

<u>Theorem:</u> Let $X_1 \ldots X_n$ be IID and let $\mu = E(X_i)$ $\sigma^2 = V(X_i)$. Then

$$\mathbb{E}(\overline{X}_n) = \mu \quad V(\overline{X}_n) = \frac{\sigma^2}{n} \quad \mathbb{E}(S_n^2) = \sigma^2$$

⇒ If X and Y are random variables, then the <u>covariance</u> and <u>correlation</u> between X and Y measure how strong the linear relation is between X and Y.

⇒ Let X and Y be random variables with means μ_X and μ_Y and standard deviations σ_X and σ_Y. Define the covariance between X and Y by

$$\boxed{Cov(X, Y) = \mathbb{E}\left((X - \mu_X)(Y - \mu_Y)\right)}$$

and the correlation by

$$\boxed{\rho = \rho_{XY} = \rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}}$$

Theorem: The covariance satisfies
$$Cov(X,Y) = \mathbb{E}(XY) - E(X)\mathbb{E}(Y)$$

The correlation satisfies
$$-1 \leq \rho(X,Y) \leq 1$$

$\Rightarrow$ If X and Y are independent, then
$$Cov(X,Y) = \rho = 0$$
$\hookrightarrow$ The converse is not true in general.

$$\boxed{V\left(\sum_i a_i X_i\right) = \sum_i a_i^2 V(X_i) + 2\sum\sum_{i<j} a_i a_j Cov(X_i X_j)}$$

$\downarrow$

{for random variables $X_1, \ldots X_n$}

★ <u>Expectation and Variance of Important Random Variables</u>

$\rightarrow$ Let random vector X be
$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$$

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_K \end{pmatrix} = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_K) \end{pmatrix}$$

$\Rightarrow$ The Covariance matrix Σ is defined as

$$\mathrm{Cov}[X] = \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_K) \\ \mathrm{Cov}(X_2 X_1) & \mathrm{Cov}(X_2 X_2) & \cdots & \mathrm{Cov}(X_2 X_K) \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}(X_K X_1) & \mathrm{Cov}(X_K X_2) & \cdots & \mathrm{Cov}(X_K X_K) \end{bmatrix}$$

$\Rightarrow$ If a is a vector and X is a random vector with mean μ and variance Σ, then $\mathbb{E}(a^T X) = a^T \mu$

&

$$\mathbb{V}(a^T X) = a^T \Sigma a$$

$\hookrightarrow$ If A is a matrix then

$$\mathbb{E}(AX) = A\mu$$

$$\mathbb{V}(AX) = A \Sigma A^T$$

★ Conditional Expectation

$\Rightarrow$ The Conditional expectation of X given $Y=y$ is

$$\mathbb{E}(X|Y=y) = \begin{cases} \sum x f_{X|Y}(x|y)\, dx & \text{discrete case} \\ \int x f_{X|Y}(x|y)\, dx & \text{Continuous case.} \end{cases}$$

Theorem: (The Rule of Iterated Expectation)

$\Rightarrow$ For random variables X and Y, assuming the expectation exist we have that

$$\mathbb{E}[\mathbb{E}(Y|X)] = E(Y)$$

$\Rightarrow$ The Conditional variance is defined as

$$\mathbb{V}(Y|X=x) = \int (y-\mu(x))^2 f(y|x)\,dy$$

where,

$$\mu(x) = \mathbb{E}(Y|X=x)$$

Theorem: For random variables X and Y.

$$\mathbb{V}(Y) = \mathbb{E}\mathbb{V}(Y/X) + \mathbb{V}\mathbb{E}(Y/X)$$

★ Moment Generating Functions.

$\Rightarrow$ The moment generating function MGF or Laplace transform of X is defined by

$$\varphi_X(t) = \mathbb{E}(e^{tX}) = \int e^{tx}\,dF(x)$$

where t varies over the real numbers.

a) ⇒ When the MGF is well defined, it can be shown that we can interchange the operations of "differentiation" and "taking expectation".

$$\psi'(0) = \left[\frac{d}{dt}\mathbb{E}e^{tX}\right]_{t=0} = \mathbb{E}\left[\frac{d}{dt}e^{tX}\right]_{t=0}$$

$$= \mathbb{E}\left[Xe^{tX}\right]_{t=0} = \mathbb{E}(X)$$

→ By taking K derivatives we conclude that $\psi^{(K)}(0) = \mathbb{E}(X^K)$.

↳ This gives us a method for computing the moments of a distribution.

⇒ Properties of the MGF.

↳ If $Y = aX + b$, then $\psi_Y(t) = e^{bt}\psi_X(at)$

↳ If $X_1 \dots X_n$ are independent and $Y = \sum_i X_i$ then $\psi_Y(t) = \prod_i \psi_i(t)$

{where ψ_i is the MGF of X_i}

⇒ Let X and Y be random variables. If $\psi_X(t) = \psi_Y(t)$ for all t in an open interval around 0, then $X \stackrel{d}{=} Y$.