3 Functions of Random Variables

Expectations

- Suppose that we perform the same experiment independently for N times and get numerical outcomes $x_1, x_2, \ldots, x_N$. Then the average outcome is $m = \frac{1}{N}\sum_i^N x_i$
- In advance of performing the experiments, by knowing the distribution, we can expect that there are $Np_X(x)$ outcomes will take value x. So the "expect average" should be $m \approx \frac{1}{N}\sum_{x\in\mathbb{R}} xNp_X(x) = \sum_{x\in\mathbb{R}} xp_X(x)$. This leads to the concept of expectation.

Definition: Expectation

A random variable $X : \Omega \to \mathbb{R}$ is said to be integrable if $\int_\Omega |X|dP < \infty$. If so, then its expectation is defined by

$$E(X) = \int_\Omega XdP = \int_\mathbb{R} xdP_X = \begin{cases} \sum_{x_i} x_i p_X(x_i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} xf_X(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

- **Remark: a random variable may not have an expectation**
  (1) Consider the Cauchy random variable with PDF $f_X(x) = \frac{1}{\pi}\frac{\lambda}{\lambda^2+x^2}$. Take $\lambda = 1$,

$$E(X) = \frac{1}{\pi}\int_{-\infty}^{\infty} x\frac{1}{1+x^2}dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} d(\ln(1+x^2)) = \frac{1}{2\pi}\ln(1+x^2)\Big|_{-\infty}^{\infty} = \infty - \infty$$

  (2) Keep tossing a coin until H appears. If H appears at the kth time, then you get ¥2^k

$$E(X) = \frac{1}{2}\times 2 + \frac{1}{4}\times 4 + \frac{1}{8}\times 8 + \cdots = \infty$$

Functions of Random Variables

- We call that a random variable $X : \Omega \to \mathbb{R}$ is a $\mathcal{F}$-measurable function on Ω. If we further consider a function $g : \mathbb{R} \to \mathbb{R}$, then this gives us a new function $g \circ X : \Omega \to \mathbb{R}$. Usually, we denote this new function as $Y = g(X)$. Note that $Y = g(X)$ may not be a random variable. To have so, we need to require that

$$\forall B \in \mathcal{B}(\mathbb{R}) : g^{-1}(B) \in \mathcal{B}(\mathbb{R}) \Rightarrow X^{-1}(g^{-1}(B)) \in \mathcal{F}$$

- If $g(X)$ is a random variable and its expectation exists, then

$$E(g(X)) = \int_\Omega g(X)dP = \int_\mathbb{R} g(x)dP_X(x) = \begin{cases} \sum_i g(x_i)p_X(x_i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x)f_X(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

- There are some useful expectations of functions
  - **Variance of** X: $var(X) = E((X - E(X))^2)$.
  - K**th Moments of** X: $E(X^k)$.
  - K**th Central Moments of** X: $E((X - E(X))^k)$.

Multiple Random Variables and Independence

- Two random variables X and Y can assign values for the same sample space Ω. Essentially, they are looking at the probability space from two different angles: one from $\sigma(X)$ and the other from $\sigma(Y)$.
- For any two Borel sets $A, B \in \mathcal{B}(\mathbb{R})$, $\{X \in A\}$ and $\{Y \in B\}$ are all events in $\mathcal{F}$ since random variable are $\mathcal{F}$-measurable. Therefore, it makes sense to consider a new measure $P_{XY} : \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) \to [0, 1]$ defined by

$$P_{XY}(A, B) = P(\{X \in A\} \cap \{Y \in B\}) = P(\{\omega \in \Omega : X \in A, Y \in B\})$$

- Recall that two events $A, B \in \mathcal{F} \subseteq 2^{\Omega}$ are independent if $P(A \cap B) = P(A)P(B)$. Therefore, we can also talk about the independence of random variables as follows.

Definition: Independence of Random Variables

Two random variables $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are said to be **independent** if for any two Borel sets $A, B \in \mathcal{B}(\mathbb{R})$, events $\{X \in A\}$ and $\{Y \in B\}$ are independent.

The case of n random variables is defined analogously.

- In fact, we have a more general definition of independence based on σ-field.

Definition: Independence of σ-Fields

Two σ-fields $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ are said to be **independent** if any two events $A \in \mathcal{G}$ and $B \in \mathcal{H}$ are independent.

- Hence, two random variables X and Y are independent iff $\sigma(X)$ and $\sigma(Y)$ are.
- We say a random variable X is independent of a σ-field $\mathcal{G} \subseteq \mathcal{F}$ if $\sigma(X)$ and $\mathcal{G}$ are independent.

- Similar to the case of CDF, by considering $A = (-\infty, x]$ and $B = (-\infty, y]$, we can define the **joint CDF** $F : \mathbb{R}^2 \to [0, 1]$ by

$$F_{XY}(x, y) = P_{XY}((-\infty, x], (-\infty, y]) = P(X \leq x, Y \leq y)$$

- Clearly, if X and Y are independent, then we have $F_{XY}(x, y) = F_X(x)F_Y(y)$. In fact, this condition is *necessary and sufficient* for independence. It essentially comes from the fact that any Borel set is measurable if and only if any set of form $(-\infty, a]$ is measurable.
- If we have an infinite sequence of random variables $X_1, X_2, \ldots$ defined on the same probability space, then we say they are independent if any finite collection of them are independent. Furthermore, if $P_{X_i} = P_{X_j}, \forall i, j$, then we say $X_1, X_2, \ldots$ are **Independent Identically Distributed (i.i.d.)**.
- Think: what if $X_1 : \Omega_1 \to \mathbb{R}$ and $X_2 : \Omega_2 \to \mathbb{R}$ are defined for different sample space? Well, we can just extend the sample space to $\Omega_1 \times \Omega_2$ by $P(A \times B) = P_1(A)P_2(B)$ and the random variables are modified to $X_1(\omega_1, -) = X_1(\omega_1)$, where "$-$" can be anything in Ω_2. Clearly, X_1 and X_2 are independent if $\Omega_1 \cap \Omega_2 = \emptyset$.

Discrete Case: Joint Probability Mass Function

- If $X : (\Omega, \mathcal{F}, P) \to \{D_X, 2^{D_X}, P_X\}$ and $Y : (\Omega, \mathcal{F}, P) \to \{D_Y, 2^{D_Y}, P_Y\}$ are discrete random variables, then we can talk about the **joint PMF** by, for any $x \in D_X, y \in D_Y$, we have
$$p_{XY}(x, y) = P(X = x, Y = y)$$
- For any two sets $A \in 2^{D_X}, B \in 2^{D_Y}$, we have
$$\begin{aligned} P_{XY}(A, B) =& P(\{\omega : X(\omega) \in A, Y(\omega) \in B\}) = P(\bigcup_{x_i \in A} \bigcup_{y_j \in B} \{\omega : X(\omega) = x_i, Y(\omega) = y_j\}) \\ =& \sum_{x_i \in A} \sum_{y_j \in B} p_{XY}(x_i, y_j) \end{aligned}$$
- For any value $x_i \in D_X$, we have
$$\begin{aligned} p_X(x_i) =& P(\{\omega : X(\omega) = x_i\}) = P(\{\omega : X(\omega) = x_i, Y(\omega) \in D_Y\}) \\ =& P(\bigcup_{y_j \in D_Y} \{\omega : X(\omega) = x_i, Y(\omega) = y_j\}) \\ =& \sum_{y_j \in D_Y} p_{XY}(x_i, y_j) \end{aligned}$$
- The PMF for new random variable (if it is) $Z = g(X, Y)$ can be computed by
$$p_Z(z) = P(Z = z) = P(\{\omega : Z(\omega) = g(X(\omega), Y(\omega)) = z\}) = \sum_{(x,y) \in g^{-1}(\{z\})} p_{XY}(x, y)$$
The expectation of $Z = g(X, Y)$ can be computed as
$$E(Z) = \sum_{z \in D_Z} z p_Z(z) = \sum_{z \in D_Z} \sum_{(x,y) \in g^{-1}(\{z\})} z p_{XY}(x, y) = \sum_{(x,y) \in D_X \times D_Y} g(x, y) p_{XY}(x, y)$$

Continuous Case: Joint Probability Density Function

- If $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ and $Y : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_Y)$ are continuous random variables, then we can talk about the **joint PDF**, which is the function $f_{XY} : \mathbb{R} \times \mathbb{R} \to [0, 1]$ such that
$$\forall B \in \mathcal{B}(\mathbb{R}^2) : P((X, Y) \in B) = \iint_B f_{XY}(x, y) dx dy$$
- Similar to the discrete case, we also have
  - $f_X(x) = \int_{\mathbb{R}} f_{XY}(x, y) dy$ and $f_Y(y) = \int_{\mathbb{R}} f_{XY}(x, y) dx$
  - $E(g(X, Y)) = \iint_{\mathbb{R}^2} g(x, y) f_{XY}(x, y) dx dy$
  - If X and Y are independent, then $f_{XY}(x, y) = f_X(x) f_Y(y)$

Computation of Functions

- **Question:**
  Given random variable $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ with PDF f_X and CDF F_X, and function $g : \mathbb{R} \to \mathbb{R}$ s.t. $Y = g(X)$ is a random variable. What are f_Y and F_Y for Y?

- **Solution:**
  $$F_Y(y) = P(g(X) \le y) = P(g(X) \in (-\infty, y]) = P(X \in g^{-1}(-\infty, y]) = \int_{g^{-1}(-\infty,y]} f(x)dx$$

- **Example:**
  Let $X \sim N(0,1)$, i.e., $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and let $Y = g(X) = X^2$. Then
  $$P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1$$
  Then differentiate to obtain
  $$f_Y(y) = 2\frac{d}{dy}\Phi(\sqrt{y}) = \frac{1}{\sqrt{y}}\Phi'(\sqrt{y}) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}y}$$

Transformation of Random Vectors

- Let $\underline{X} = (X_1, X_2, \ldots, X_n)^{\mathrm{T}}$ be a random vector, where $X_i : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_{X_i})$. It can also be treated as a single random variable $\underline{X} : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_{\underline{X}})$.

- Let us consider a transformation $g : \mathbb{R}^n \to \mathbb{R}^n$. This gives us a new random vector $\underline{Y} = g(\underline{X})$, where $Y_1 = g_1(X_1, X_2, \ldots, X_n), \ldots, Y_n = g_n(X_1, X_2, \ldots, X_n)$.

- Suppose that the joint PDF $f_{\underline{X}}$ is given. How to determine the joint PDF $f_{\underline{Y}}$ of $\underline{Y}$?

- Here we assume that the transformation is invertible, i.e., g^{-1} exists, and g^{-1} is continuous and has continuous derivatives. Then the solution is as follows.

  Consider any $C \in \mathcal{B}(\mathbb{R}^n)$. Then
  $$P(\underline{Y} \in C) = P(\{\omega : g(\underline{X}(\omega)) \in C\}) = P(g(\underline{X}) \in C) = P(\underline{X} \in B),$$
  where $B = \{\underline{x} \in \mathbb{R}^n : g(\underline{x}) \in C\}$. Since g^{-1} exists, we assume $\underline{X} = h(\underline{Y})$. Then
  $$P(\underline{Y} \in C) = P(\underline{X} \in B) = \int_B f_{\underline{X}}(\underline{x})d\underline{x} = \int_C f_{\underline{X}}(h(\underline{y}))|\det(J(h(\underline{y})))|d\underline{y}$$
  where $J(h(\underline{y}))$ is the Jacobian matirx
  $$J(h(\underline{y})) = \begin{bmatrix} \frac{\partial h_1(y)}{\partial y_1} & \frac{\partial h_1(y)}{\partial y_2} & \cdots & \frac{\partial h_1(y)}{\partial y_n} \\ \frac{\partial h_2(y)}{\partial y_1} & \frac{\partial h_2(y)}{\partial y_2} & \cdots & \frac{\partial h_2(y)}{\partial y_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial h_n(y)}{\partial y_1} & \frac{\partial h_n(y)}{\partial y_2} & \cdots & \frac{\partial h_n(y)}{\partial y_n} \end{bmatrix}, h = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix}$$
  Since the above is true for any $C \in \mathcal{B}(\mathbb{R}^n)$, we conclude that
  $$f_{\underline{Y}}(\underline{y}) = f_{\underline{X}}(h(\underline{y}))\det(J(h(\underline{y})))$$

Independent v.s. Uncorrelated

- In some problems, we are interested in whether or not random variables X and Y have some **linear relation**. For this requirement, we can use the **covariance** of X and Y:

$$cov(X,Y) = E((X-E(X))(Y-E(Y))) = E(XY) - E(X)E(Y)$$

- The **correlation** of X and Y is defined as

$$\rho(X,Y) = E\left[\left(\frac{E(X-E(X))}{\sqrt{var(X)}}\right)\left(\frac{E(Y-E(Y))}{\sqrt{var(Y)}}\right)\right] = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}$$

  We say X and Y are **uncorrelated** if $\rho(X,Y)=0$, i.e., $cov(X,Y)=0$.

Properties of Expectations

- $\boxed{E(aX+bY) = aE(X)+bE(Y)}$.
  Proof: Only consider the discrete case. Let $A_x = \{X=x\}$ and $B_y = \{Y=y\}$. Then

$$(\star) = E(\sum_{x,y}(ax+by)\mathbf{1}_{A_x\cap B_y}) = \sum_{x,y}(ax+by)P(A_x\cap B_y) = \sum_{x,y} axP(A_x\cap B_y) + \sum_{x,y} byP(A_x\cap B_y)$$

  However, we have

$$\sum_y P(A_x\cap B_y) = P\left(A_x\cap(\cup_y B_y)\right) = P(A_x\cap\Omega) = P(A_x) \text{ and } \sum_x P(A_x\cap B_y) = P(B_y)$$

  Therefore, we have

$$(\star) = \sum_x ax\sum_y P(A_x\cap B_y) + \sum_y by\sum_x P(A_x\cap B_y) = a\sum_x xP(A_x) + b\sum_y yP(B_y) = aE(X)+bE(Y)$$

- If X and Y are *independent*, then $\boxed{E(g(X)h(Y)) = E(g(X))E(h(Y))}$.
  Proof: Still for the discrete case.

$$E(g(X)h(Y)) = \sum_{x,y} g(x)h(y)P(A_x\cap B_y) = \sum_{x,y} g(x)h(y)P(A_x)P(B_y) = \sum_x g(x)P(A_x)\sum_y h(y)P(B_y)$$

- (1) $var(aX) = a^2 var(X)$;
  (2) If X and Y are uncorrelated, the $var(X+Y) = var(X)+var(Y)$.

- The above shows that independent implies uncorrelated. However, **the converse is not necessarily true** because uncorrelated just means that X and Y do not have *linear* relation. For example, consider X and Y with the following joint PMF

$$P_{XY}(1,1) = P_{XY}(1,-1) = 0.25 \text{ and } P_{XY}(-1,0) = 0.5$$

  - X and Y are uncorrelated: $cov(X,Y) = E(XY) - E(X)E(Y) = 0$
  - X and Y are dependent, however: $P(X=1,Y=1) = \frac{1}{4} \neq P(X=1)P(Y=1) = \frac{1}{8}$