# Probability by Yiding 2024.11.17

## Basic Definitions :

1. $\Omega$ universal set(sample space);
$A$ event; $\omega$ element; $A \subseteq \Omega$; $\omega \in \Omega$
2. Probability Measure(Function):
(1).$\mathbf{P}(A) = \frac{|A|}{|\Omega|} \geq 0 \ \forall A$;
(2). $\mathbf{P}(\Omega) = 1$;
(3). for $A_i \cap A_j = \emptyset, \mathbf{P}(A_i \cup A_j) = \mathbf{P}(Ai) + \mathbf{P}(Aj)$
3. n pick k (order, replacement) $= n^k, (1,1)$; $\binom{n}{k}, (0,0)$; $\frac{n!}{(n-k)!}, (1,0)$; $\binom{n+k-1}{k}, (0,1)$;

4.**Inclusion-Exclusion Principle** Let $n \in \mathbb{N}$ and $A_i, i \in \{1,\ldots,n\}$ be events. Then$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{1\leq i_1<i_2<\cdots<i_k\leq n} P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right)$

5. $Min(\mathbb{P}(A), \mathbb{P}(B)) \geq \mathbb{P}(A \cap B)$ **Upper bound for $\cap$**
6. $Max(\mathbb{P}(A), \mathbb{P}(B)) \leq \mathbb{P}(A \cup B)$ **Lower bound for $\cup$**
7. If $B \subseteq A$, then$B \cap A = B$, $P(B \cap A) = P(B)$
8. Let $A, B \in \Omega$ be events with $A \subseteq B$. Then $P(A) \leq P(B)$.

## Conditional Probability :

1. $P(A \cap B) = P(A \mid B) \cdot P(B)$
2.**The multiplication rule**:$P(E_1E_2E_3\cdots E_n) = P(E_1)P(E_2 \mid E_1)P(E_3 \mid E_1E_2)\cdots P(E_n \mid E_1\cdots E_{n-1})$
3.**The Law of Total Probability** $P(B) = P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_n)P(A_n)$; $P(B) = P(B \mid A)P(A) + P(B \mid A^c)P(A^c)$
4.**Bayes' Formula**: $P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$
5.**Bayes' Theorem** Let our sample space $\Omega$ be partitioned into disjoint events $A_k$, $k = 1, 2, \ldots, n$. Then for $1 \leq j \leq n$,$P(A_j \mid B) = \frac{P(B|A_j)P(A_j)}{\sum_{k=1}^n P(B|A_k)P(A_k)}$, (Use 3. to express $P(B)$)
6.**Independence of A and B** If $P(A \cap B) = P(A)P(B)$ it is just $P(A) = P(A \mid B)$

## Random Variable :

1. **RV - Random Variable**: sth in $\Omega$, the outcome of some experiment. Some we can list all the RV, others we can't. Denote as $X$.
2. **CDF - Cumulative Distribution Function**: Let X be RV, $F_X(x) := P(X \leq x)$.
3. Let $X$ be a random variable and $F_X$ its cdf. Then $F_X$ is non-decreasing; $F_X$ is right-continuous; $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$. Furthermore, any function that fulfills the three properties above is a cdf.
4. **Discrete RV - Probability Mass Function PMF**: $p(x_j) = P(X = x_j)$
5. **Continuous RV - Probability Density Function PDF**: $P(X \in B) = \int_B f_X(x)\,dx$; $F_X(x) = \int_{-\infty}^x f_X(t)\,dt$; $f_X(x) = F'_X(x)$; $P(X = x) = 0$; Always True : $f_X(x) \geq 0$ for all $x \in \mathbb{R}$; $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

## Discrete RV :

1. $X \sim unif(A)$, $P(X = x) = \frac{1}{|A|}$, $\forall x \in A$
2. $X \sim Bernoulli(p)$, $P(X = 1) = p, P(X = 0) = 1 - p$, $E[X] = p$
3. $X \sim Bin(n,p)$ *or* $Binom(n,p)$, $P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$, $E[X] = np, Var(X) = np(1-p)$
4. $X \sim Geom(p)$, $P(X = k) = (1-p)^{k-1}p$, $\forall k \in 1, 2, ..., ...,$ $E[X] = \frac{1}{p}$, $Var(x) = \frac{1-p}{p^2}$
5. $X \sim Poisson(\lambda)$, $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$, $\forall k \in 0, 1, 2, ..., ..., E(x) = \lambda, Var(x) = \lambda$
6. Relation of the Poisson and Binomial distribution: If $X \sim Bin(n,p)$ and $n$ is large and $p$ is small, then $X \sim Poisson(\lambda)$ with $\lambda = np$.

## Continuous RV :

1. $X \sim Unif([a,b])$, $f(x) = \mathbf{1}_{[a,b]}(x)\frac{1}{b-a}$, $F_X(x) = \frac{x-a}{b-a}$,$E[X] = \frac{a+b}{2}$, $E(X) = \frac{1}{2}$, $Var(X) = \frac{1}{12}$
2. $X \sim Exp(\lambda)$, $f(x) = \mathbf{1}_{[0,+\infty)}(x)\lambda e^{-x\lambda}$, $F_X(x) = (1 - e^{-x\lambda})$, Memoryless: $\mathbb{P}(X > t + s | X > s) = \mathbb{P}(X > t)$., $E(X) = \frac{1}{\lambda}$

## Independence :

1. $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$
2. $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$
3. **Expectation of Independent RVs**: $E[XY] = E[X]E[Y]$
4. **Variance of Independent RVs**: $Var(X+Y) = Var(X) + Var(Y)$, $Var(aX + bY) = a^2Var(X) + b^2Var(Y)$

## Expectation :

1. **Expectation**: $E[X] = \sum_x x \cdot p(x)$ *or* $\int_{-\infty}^{\infty} x \cdot f(x)\,dx$
$\mathbb{E}[X] = \sum_{x_i} xi \cdot P(X = xi)$
$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)dx.$
2. **median**: $F_X(x) = \frac{1}{2}$
3. **mode**: $f_X(x)$ Maximum
4. For $E[X] < \infty$, $E[aX + Y + b] = aE[X]E[Y] + b$, $E[X + Y] = E[X] + E[Y]$ E is linear
5. **Sample Mean for iid RVs $X_i$ that $E(X) = \mu$**: $\overline{X} = \sum_{i=1}^n \frac{X_i}{n}, E(\overline{X}) = \mu$
6.If $X, Y$ are discrete with joint pmf $p_{X,Y}$, then: $\mathbb{E}[g(X,Y)] = \sum_x \sum_y g(x,y)p_{X,Y}(x,y)$.
If $X, Y$ are continuous with joint pdf $f_{X,Y}$, then: $\mathbb{E}[g(X,Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x,y)f_{X,Y}(x,y)\,dy\,dx$.
7. **Conditional Probability**: $E[g(X)|A] = \sum_{X_i} g(xi)P(X = x_i|A)$

## Variance :

1. **Variance**: $Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$
2. $Var(aX + b) = a^2Var(X)$

## Transformation :

1. **Linear Transformation**$Y := aX + b$: $f_Y(y) = \frac{1}{|a|}f_X\left(\frac{y-b}{a}\right)$
2. **Monotone Transformation**: For strictly monotone function $Y := g(x)$, $f_Y(y) = f_X\left(g^{-1}(y)\right) \cdot \left|\left(g^{-1}\right)'(y)\right|$
3. **the absolute sign is given by the monotonic decease and negativity of a, Be careful about the range(domain) of Y**

## Normal Distribution :

1. **Standard Normal Distribution**: $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, $X \sim N(\mu, \sigma^2)$, where $\mu = 0, \sigma = 1$ 2.
**Normal Distribution**: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

## Joint Distributed RV :

**Marginal Density**:
$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)dy$
$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)dx$
**Marginal pmf**:
$P_X(x) = \sum_y p_{X,Y}(x,y)$
$P_Y(y) = \sum_x p_{X,Y}(x,y)$

**Independence of RV iff**:
$P_{X,Y}(x,y) = P_X(x) \cdot P_Y(y)$
$f_{X,Y}(x,y) = f_X(x) \cdot f_y(y)$

**Sum(Convolution) of Independent RV** :
1. $P_{X+Y}(k) = \sum_y P_X(k-y) \cdot P_Y(y)$
2. $f_{X+Y}(k) = \int_{\mathbb{R}} f_X(k-y) \cdot f_Y(y)dy$
3. Let $X \sim \text{Poisson}(\lambda_X)$ and $Y \sim \text{Poisson}(\lambda_Y)$ be independent. Then Z := X + Y has distribution $Z \sim \text{Poisson}(\lambda_X + \lambda_Y)$.
4. Let $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$ be independent. Then $X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$ .

**Covariance and Correlation** :
1. **Covariance:** $\text{Cov}(X,Y) = E[XY] - E[X]E[Y] = E(X - E[X])(Y - E[Y])$
2. **Covariance Property:** $\text{Cov}(aX + bY, Z) = a\,\text{Cov}(X,Z) + b\,\text{Cov}(Y,Z)$ for all $a, b \in \mathbb{R}$.
3. $(Cov(X,Y))2 \leq Var(X)Var(Y)$
4. **Correlation:** $\text{Cor}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \in [-1,1]$
5. **More Variance**:
$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
6. if $|Cor(X,Y)| = 1$, $tX - Y = c$ where $c$ and is a constant and $t = \frac{Cov(X,Y)}{Var(X)}$, this means $X$ and $Y$ are highly dependent where $Y$ is just the linear transformation of $X$ 7. Two normally distributed random variables $X, Y$ are independent if and only if $Cov(X,Y) = 0$.

**Maximum and Minimum of RVs** :
$X_i$ be iid RVs
The probability of Max $\leq z$ means $\forall X_i, X_i \leq z$
The probability of Min $\leq z$ means **at least one** $X_i, X_i \leq z$ – Use $(1-P)$
1. $Ma_n := \max_{i \in \{1,\ldots,n\}}(X_i)$ :
$F_{Ma_n}(z) = \mathbb{P}(Max \leq z) = \prod_{i=1}^n F_{X_i}(z) = (F_{X_i}(z))^n$
2. $Mi_n := \min_{i \in \{1,\ldots,n\}}(X_i)$ :
$F_{Mi_n}(z) = \mathbb{P}(Min \leq z) = 1 - \prod_{i=1}^n 1 - F_{X_i}(z) = 1 - (1 - F_{X_i}(z))^n$
3. Let $X_i$ be independent $Exp(\lambda_i)$ distributed RVs. Then $Y := min_{1 \leq i \leq n} X_i$ has distribution $Y \sim Exp(\sum_{i=1}^n \lambda_i)$.

**Important Summations** :
1. $\sum_{k=0}^n k^2 = \frac{n(n+1)(2n+1)}{6}$
2. $\sum_{k=0}^n k^3 = \left(\frac{n(n+1)}{2}\right)^2$
3. $e^\lambda = \sum_{k=0}^n \lambda^k / k!$
4. $sin(x) = \sum_{k=0}^n (-1)^k x^{2k+1}/(2k+1)!$
5. $cos(x) = \sum_{k=0}^n (-1)^k x^{2k}/(2k)!$

**(Central) Limit Theorem** :
1. **Markov's inequality** : $P(X \geq a) \leq \frac{E[X]}{a}$
2. **Chebyshev's inequality** : $P(|X - E[X]| \geq a) \leq Var(X)/a^2$
3. **CLT- Central Limit Theorem**:
$Let X_1, \ldots, X_n$ be iid random variables with $\mathbb{E}[X_1] = \mu \in \mathbb{R}$ and $\text{Var}(X_1) = \sigma^2 > 0$. We define $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$. Then$\mathbb{P}\left(\frac{\overline{X}_n - \mathbb{E}[\overline{X}_n]}{\sqrt{\text{Var}(\overline{X}_n)}} \leq x\right) \xrightarrow{n\to\infty} \Phi(x)$.
4. **Other form of CLT**:
$\mathbb{P}\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x\right) \approx \Phi(x)$
5. **Weak LLN Law of Large Number**: Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables with finite variance. Then for any $a > 0$, we have
$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}[X_1]\right| \geq a\right) = 0$.

**Moment Generating Function** :
1. Let $X$ be a random variable. The moment generating function (MGF) of $X$, denoted by $M_X(t)$, is defined as:
$M_X(t) = E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} f_X(x)dx$ or $\sum_x e^{tx} \cdot P_X(x)$
2. $E(X^n) = M_X^{(n)}(0)$
3. $X \sim Bernoulli(p)$, $M_X(t) = 1 - p + pe^t$
4. $X \sim Exp(\lambda)$, $M_X(t) = \frac{\lambda}{1-\lambda}$
5. $X \sim N(\mu, \sigma^2)$, $M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$
6. $M_{X_1+X_2++X_n}(t) = M_{X_1}(t)M_{X_2}(t)...M_{X_n}(t)$

**Tips** :
1. **METHOD: When dealing with At Least problem, To find $\mathbb{P}(A)$ Better to use $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$**
2. **iid = independent identically distributed**
3. **METHOD: Fare 6D**: $E(X) = 3.5, Var(x) = \frac{35}{12}$
4. continuous(like time) and memoryless –> Exp($\lambda$)
5. **Integral of pdf is always 1, can use for calculation**

**Example** :
1. **- Cov(X,Y) = 0, but X, Y are not independent /***/ - Var(X+Y) = Var(X) + Var(y), But X,Y are not independent**:
$X \sim unif[-1,1], Y = X^2, Cov(XY) = E[XY] - E[X]E[Y]$
$f_X(x) = \frac{1}{2} \cdot \mathbf{1}_{[-1,1]}(x)$, $E[XY] = E[X^3] = \int_{-1}^1 \frac{1}{2}x^3 dx = 0$
$Similarly E(X) = 0$, therefore $Cov(XY) = 0$ but $Y = X^2$ which means they are not independent