(Lecture 4) properties $\begin{cases} 0 \le p(x) \le 1, & x = 0, 1, \cdots \\ \sum_{x=0}^{\infty} p(x) = 1 \end{cases}$

A. Discrete RV / prob distribution

A1. Binomial prob distribution (二项分布)

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \cdots, n$$

$X \sim B(n,p)$, $0 \le p \le 1$, $n > 0$

$C_n^k = \frac{n!}{k! \times (n-k)!}$ $\boxed{\begin{matrix}\text{成功 } p \\ \text{失败 } 1-p\end{matrix}}$ pmf (prob mass function) 离散RV

$E(X) = np \quad Var(X) = np(1-p)$

A2. Poisson distribution (泊松分布)

$$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \lambda > 0, \; k \in \{0, 1, \cdots\}$$

$X \sim Poisson(\lambda)$, $E(X) = Var(X) = \lambda$

$X \sim Poisson(\lambda_1)$, $Y \sim Poisson(\lambda_2)$

$X + Y \sim Poisson(\lambda_1 + \lambda_2)$

A3. Geometric prob distribution (几何分布)

$$P(X=k) = (1-p)^{k-1} \cdot p, \quad k = 1, 2, \cdots$$

$X \sim G(p)$, $E(X) = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$

(Lecture 5)

B. Continuous RV / prob distribution

(a real-valued function $f: R \to R$ satisfying $f(x) \ge 0$, $x \in R$, and $\int_R f(x)dx = 1$)

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(y)dy$$

cdf (cumulative distribution function) 连续RV $\Downarrow$

pdf (prob density function)

B1. Normal (Gaussian) prob distribution (正态/高斯) $X \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

$E(X) = \mu$, $Var(X) = \sigma^2$

B2. Uniform distribution on $(a,b)$ (均匀分布)

$X \sim U([a,b])$. $E(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & x \notin [a,b] \end{cases}$$

B3. Gamma prob distribution (gamma分布)

$$X \sim \Gamma(\alpha, \beta), \; f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

$\alpha > 0$, $\beta > 0$

Gamma function $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$

$\Gamma(1) = 1$, $\Gamma(a+1) = a\Gamma(a)$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

$\Gamma(a) = (a-1)!$

B4. Exponential distribution (指数分布)

$X \sim Exp(\lambda)$, $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$, $x \in [0, +\infty)$, $\lambda > 0$.

$F_X(x) = 1 - e^{-\lambda x}$, $x \in [0, +\infty)$, $E(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$

B5. $\chi^2$ prob distribution (卡方分布)

$X \sim N(0,1)$, $Y = X^2$, $F_Y(y) = P(Y \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = \int_0^y \frac{1}{\sqrt{2\pi u}} e^{-u/2} du$

$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}$, $Y \sim \Gamma(\frac{1}{2}, 2) = \chi^2(1)$

$X \sim \chi^2(k)$. $f_k(x) = \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} x^{k/2 - 1} e^{-x/2}$, $x \in [0, +\infty)$

B6. Rayleigh prob distribution (瑞利分布)

$X \sim Rayleigh(\sigma)$, $\sigma > 0$, $f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$, $x \in [0, +\infty)$, when $\sigma = 1$, $f(x) = xe^{-\frac{x^2}{2}}$

$f(x) \ge 0$, $\int f(x)dx = 1$

$X_1, X_2 \sim N(0,1)$ independent

$R = \sqrt{X_1^2 + X_2^2} \sim Rayleigh(1)$

(Lecture 6) Random Vectors, prob distribution, Multidimensional cdf/pdf

$X = (X_1, X_2, \cdots, X_n): \Omega \to R^n$, where each $X_i: \Omega \to \mathbb{R}$ is random variable

$F_X(x_1, \cdots, x_n) = P(\omega: X_1(\omega) \le x_1, \cdots, X_n(\omega) \le x_n)$

$F_{X_i}(x_i) = P(X_i \le x_i) := P(X_1 \in R, \cdots, X_i \le x_i, \cdots, X_n \in R) = F_X(\infty, \cdots, x_i, \cdots \infty)$

Definition 1 $X_1, \cdots X_n$ are independent iff $F_X(x_1, x_2, \cdots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$ or $f(x_1, \cdots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$

Definition 2 $X = (X_1, \cdots, X_n)$ jointly continuous if there is a function $f: R^n \to R$ s.t.

1) $f(x) \ge 0$, all $x \in R^n$ 2) $\int_{R^n} f(x)dx = 1$, 3) $P(\omega: X(\omega) \in B) = \int_B f(x)dx$, all Borel set $B \in \mathcal{B}(\mathbb{R}^n)$

$F_X(x_1, \cdots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(y_1, \cdots, y_n) dy_n \cdots dy_1$ joint pdf of vector X

Definition 3 $X = (X_1, \cdots, X_n) \to$ jointly continuous Gaussian distribution

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det R}} \exp\left(-\frac{1}{2}(x-\mu)^T R^{-1} (x-\mu)\right), \text{ where } x = (x_1, \cdots, x_n) \in R^n$$

vector $\mu \in R^n$, $R$, $n \times n$ positive definite matrix

Simple example $X_i \sim N(\mu_i, \sigma_i^2)$ independent RV.

$$f_X(x) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) = \frac{1}{\sqrt{(2\pi)^n \det R}} \exp\left(-\frac{1}{2}(x-\mu)^T R^{-1}(x-\mu)\right)$$

$\mu = (\mu_1, \cdots, \mu_n)$, $R$ diagonal matrix with element $\sigma_i^2$

(Lecture 7) Transformation, Conditional densities

$$X = (X_1, \cdots, X_n) \underset{\varphi^{-1}}{\overset{\varphi}{\rightleftarrows}} Y = (Y_1, \cdots, Y_n) \quad \varphi = (\varphi_1, \varphi_2, \cdots, \varphi_n) \quad \frac{Y = \varphi(X)}{X = \varphi^{-1}(Y)}$$

Given joint pdf $f_X(x) \xrightarrow{?} f_Y(y)$ inverse transformation

$$P_Y(B) = P(Y = \varphi(X) \in B) = P(\varphi(X) \in B) = P(X \in \varphi^{-1}(B))$$

$$= \int_{\varphi^{-1}(B)} f_X(x)dx = \int_B f_X(\varphi^{-1}(y)) |\det D\varphi^{-1}(y)| dy$$

$x = \varphi^{-1}(y)$, $dx = |\det D\varphi^{-1}(y)| dy$

$$\Rightarrow f_Y(y) = f_X(\varphi^{-1}(y)) |\det D\varphi^{-1}(y)|$$

Jacobi matrix

$$\begin{pmatrix} \frac{\partial \varphi_1^{-1}(y_1, \cdots, y_n)}{\partial y_1} & \cdots & \frac{\partial \varphi_1^{-1}}{\partial y_n} \\ \vdots & & \\ \frac{\partial \varphi_n^{-1}}{\partial y_1} & \cdots & \frac{\partial \varphi_n^{-1}}{\partial y_n} \end{pmatrix}$$

Example $X = (X_1, X_2)$, $X_i \sim N(0,1)$ independent

polar coordinate, transformation

$(X_1, X_2) \longrightarrow (R, \Theta)$

$r = \varphi_1(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$, $\theta = \varphi_2(x_1, x_2) = \arctan \frac{x_2}{x_1}$

$x_1 = \varphi_1^{-1}(r, \theta) = r\cos\theta$, $x_2 = \varphi_2^{-1}(r, \theta) = r\sin\theta$

$$|\det D\varphi^{-1}(r,\theta)| = \left\| \begin{matrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{matrix} \right\| = |r\cos^2\theta + r\sin^2\theta| = r$$

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{x_1^2 + x_2^2}{2}} \Longrightarrow f_{R,\Theta}(r, \theta) = \frac{1}{2\pi} e^{-\frac{r^2}{2}} \cdot r$$

$(r > 0, \; 0 \le \theta < 2\pi)$

**Definition 1** conditional prob density

$$f_{X_1|X_2=x_2}(x_1,x_2) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_2}(x_2)}$$

(1-dimensional pdf) $\int f_{X_1|X_2=x_2}(x_1,x_2)\,dx_1 = 1$ ← 固定数

$$P(X_1\in B_1|X_2=x_2) = \int_B f_{X_1|X_2=x_2}(x_1,x_2)\,dx_1$$

$$g_1(x_2) = E(X_1|X_2=x_2) = \int x_1 f_{X_1|X_2=x_2}(x_1,x_2)\,dx_1$$

**Example** X, Y jointly continuous vector, independent, pdf: $f_X(x)$, $f_Y(y)$ define $Z = X+Y$. pdf of Z?

$$F_Z(z) = P(X+Y\le z) = \iint_{x+y\le z} f_{XY}(x,y)\,dx\,dy \underset{(x=z-y)}{=} \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{z-y} f_X(x)\,dx\right) f_Y(y)\,dy$$

$$= \int_{-\infty}^{+\infty} F_X(z-y)f_Y(y)\,dy \xRightarrow[\text{diff } z]{} f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)\,dy$$

## (Lecture 8) Expectation, Variance, properties

X discrete RV. $EX = \sum_{i=1}^{\infty} x_i P(\omega: X(\omega)=x_i) = \sum_{i=1}^{\infty} x_i p(x_i)$

X continuous RV. $EX = \int_R x f_X(x)\,dx$

**Example1** a) $X\sim B(n,p)$, $EX = \sum_{i=0}^{\infty} i\binom{n}{i}p^i(1-p)^{n-i} = np$

b) $X\sim Poission(\lambda)$, $EX = \sum_{i=0}^{\infty} i\frac{\lambda^i}{i!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{i=1}^{\infty}\frac{\lambda^{i-1}}{(i-1)!}$ $(k=i-1)$

$$=\lambda e^{-\lambda}\underbrace{\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}}_{e^{\lambda}} = \lambda\cdot e^{-\lambda}e^{\lambda} = \lambda$$

c) $X\sim N(\mu,\sigma^2)$, $EX = \int_{-\infty}^{+\infty} x\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx \underset{(z=\frac{x-\mu}{\sigma})}{=} \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}(\mu+\sigma z)e^{-\frac{1}{2}z^2}\,dz$

$$=\mu\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\,dz + \sigma\underbrace{\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}ze^{-\frac{1}{2}z^2}\,dz}_{\text{odd } =0} = \mu\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\,dz + 0\sigma = 1\cdot\mu = \mu$$

**Definition1** $\phi:\mathbb{R}\to\mathbb{R}$ measurable map/function if $\phi^{-1}(B)\in\mathcal{B}(\mathbb{R})$ for all $B\in\mathcal{B}(\mathbb{R})$

X. RV. $\phi$ measurable function, $Y=\phi(X) \Rightarrow$ Y is again RV on $(\Omega,\mathcal{F},P)$

**Properties of expectation**

1. $E(\phi(x)) = \int\phi(x)f_X(x)\,dx$ 2. $E(aX+bY) = aE(X)+bE(Y)$

3. X, Y independent, $E(\phi(X)\varphi(Y)) = E(\phi(X))\cdot E(\varphi(Y))$

**Definition2** any $n=1,2,\dots$, $\mu_n = E(X^n)$ called n-th moment of X

**Definition3** $Var(X) = \sigma^2 = \sigma_X^2 = E[(X-\mu_X)^2]$

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

## (Lecture 1) Probability space $(\Omega,\mathcal{F},P)$

countable
- finite countable
- infinite countable

**Definition1** Let $\Omega\neq\emptyset$. Then a collection $\mathcal{F}$ of subsets of $\Omega$ is called a σ-algebra if

1) $\emptyset\in\mathcal{F}$ / $\Omega\in\mathcal{F}$

2) if $A\in\mathcal{F}$, then $A^c = \Omega\setminus A\in\mathcal{F}$

3) if $A_i\in\mathcal{F}$, $i=1,2,\dots$, then $\bigcup_{i=1}^{\infty}A_i\in\mathcal{F}$ ($\bigcup_{i=1}^{n}$, $\mathcal{F}$ is called an algebra)

or $\bigcap_{i=1}^{\infty}A_i\in\mathcal{F}$ $\left(\bigcap_{i=1}^{\infty}A_i = \left(\bigcup_{i=1}^{\infty}A_i^c\right)^c\right)$

De Morgan's Law $(A\cap B)^c = A^c\cup B^c$; $(A\cup B)^c = A^c\cap B^c$

**Definition2** (Kolmogoroff) Let $\Omega\neq\emptyset$ and $\mathcal{F}$ is a σ-algebra on $\Omega$.
A set function $P:\mathcal{F}\to[0,1]$ is called a probability measure on $\mathcal{F}$ if it satisfies the following axioms: 概率测度

1) $P(\Omega)=1$ 2) if $A_i\in\mathcal{F}$, pairwise disjoint, then $P(\bigcup_{i=1}^{\infty}A_i) = \sum_{i=1}^{\infty}P(A_i)$

$P(A)\ge 0$ for all A (replace $\bigcup_{i=1}^{\infty}A_i$ to $\bigcup_{i=1}^{n}A_i$, P called additive function)

**Properties of probability function**

a) $P(A^c) = 1-P(A)$ b) $A\subset B$. $P(A)\le P(B)$ c) $P(A\cup B) = P(A)+P(B)-P(A\cap B)$

d) μ - prob function, $A_1\subset A_2\subset A_3\subset\cdots$, $A=\bigcup_{n=1}^{\infty}A_n$

$\lim_{n\to\infty}\mu(A_n) = \mu(A)$. Define $B_1=A_1$, $B_n = A_n\setminus A_{n-1}$ (disjoint). $A=\bigcup_{n=1}^{\infty}A_n = \bigcup_{n=1}^{\infty}B_n$

$(\Omega, \mathcal{F}, P)$ 描述一个完整概率模型

sample space, σ-algebre, probability measure

## (Lecture 2) probability space $(\Omega,\mathcal{F},P)$: continued

**Definition1** $P(B|A) = \frac{P(A\cap B)}{P(A)}$ (conditional prob)

**Definition2** Assume $A\cap B\neq\emptyset$. A, B independent iff $P(A\cap B) = P(A)\cdot P(B)$

**Definition3** $A_1,\dots,A_n$, partition of $\Omega$, $\begin{cases} A_i\cap A_j=\emptyset \\ \bigcap_{i=1}^{n}A_i = \Omega\end{cases}$

for any event B. $P(B) = \sum_{i=1}^{n}P(B|A_i)P(A_i)$ (formula of total probability)

$$P(A_j|B) = \frac{P(B|A_j)\cdot P(A_j)}{\sum_{i=1}^{n}P(B|A_i)P(A_i)}$$

**Definition4** two σ-algebras $\mathcal{G}$ and $\mathcal{F}$ are said to be independent iff $P(A\cap B) = P(A)\cdot P(B)$, $A\in\mathcal{G}$, $B\in\mathcal{F}$

$A\in\mathcal{F}$, $\mathcal{F}_A = \{\emptyset,\Omega,A,A^c\}$ ← σ-algebra generated by the event A

**Proposition** A, B independent iff $\mathcal{F}_A$, $\mathcal{F}_B$ independent

**Proposition** $\mathcal{A}$ any collection of subsets from $\Omega$, There exist the smallest σ-algebra which contains all elements from $\mathcal{A}$. σ-algebre ← σ(A)

**Definition5** $\mathcal{A}$ be the collection of all open sets in $R^n$, σ(A) is the smallest σ-algebra contains all open sets in $R^n$, is called Borel σ-algebra on $\mathbb{R}^n$ or σ-algebra of Borel sets in $\mathbb{R}^2$
$\Downarrow$
$\mathcal{B}(\mathbb{R}^n)$

## (Lecture 3) Probability measure on Borel Sets, random variable

**Definition1** A map $X:\Omega\to\mathbb{R}^n$ is said to be a random variable if the inverse image of any Borel set $B\in\mathcal{B}(R^n)$ is an event in $\mathcal{F}$

$$X^{-1}(B) = \{\omega: X(\omega)\in B\}\in\mathcal{F} \text{ for any } B\in\mathcal{B}(R^n)$$

$$X:(\Omega,\mathcal{F})\longrightarrow(\mathbb{R},\mathcal{B}(\mathbb{R}))$$

## (Lecture 1) Probability space $(\Omega, \mathcal{F}, P)$

**Definition 1** Let $\Omega \neq \phi$. Then a collection $\mathcal{F}$ of subsets of $\Omega$ is called a $\sigma$-algebra if

1) $\phi \in \mathcal{F}$ / $\Omega \in \mathcal{F}$
2) if $A \in \mathcal{F}$, then $A^c = \Omega \backslash A \in \mathcal{F}$
3) if $A_i \in \mathcal{F}$, $i = 1, 2, \cdots$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ ($\bigcup_{i=1}^{n}$, $\mathcal{F}$ is called an algebra)

or $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$ $\left( \bigcap_{i=1}^{\infty} A_i = \left( \bigcup_{i=1}^{\infty} A_i^c \right)^c \right)$

De Morgan's Law $(A \cap B)^c = A^c \cup B^c$; $(A \cup B)^c = A^c \cap B^c$

**Definition 2** (Kolmogoroff) Let $\Omega \neq \phi$ and $\mathcal{F}$ is a $\sigma$-algebra on $\Omega$. A set function $P: \mathcal{F} \to [0,1]$ is called a probability measure on $\mathcal{F}$ 概率测度 if it satisfies the following axioms:

1) $P(\Omega) = 1$ 2) if $A_i \in \mathcal{F}$, pairwise disjoint, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

$P(A) \geq 0$ for all $A$ (replace $\bigcup_{i=1}^{\infty} A_i$ to $\bigcup_{i=1}^{n} A_i$, $P$ called additive function)

**Properties of probability function**

a) $P(A^c) = 1 - P(A)$ b) $A \subset B$, $P(A) \leq P(B)$ c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

d) $\mu$ - prob function, $A_1 \subset A_2 \subset A_3 \subset \cdots$, $A = \bigcup_{n=1}^{\infty} A_n$

$\lim_{n \to \infty} \mu(A_n) = \mu(A)$, Define $B_1 = A_1$, $B_n = A_n \backslash A_{n-1}$ (disjoint), $A = \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$

## (Lecture 2) probability space $(\Omega, \mathcal{F}, P)$: continued

**Definition 1** $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$ (conditional prob)

**Definition 2** Assume $A \cap B \neq \phi$. $A, B$ independent iff

$$P(A \cap B) = P(A) \cdot P(B)$$

**Definition 3** $A_1, \cdots, A_n$, partition of $\Omega$, $\begin{cases} A_i \cap A_j = \phi \\ \bigcap_{i=1}^{n} A_i = \Omega \end{cases}$

for any event $B$. $P(B) = \sum_{i=1}^{n} P(B|A_i) P(A_i)$ (formula of total probability)

$$P(A_j|B) = \frac{P(B|A_j) \cdot P(A_j)}{\sum_{i=1}^{n} P(B|A_i) P(A_i)}$$

**Definition 4** two $\sigma$-algebras $\mathcal{G}$ and $\mathcal{F}$ are said to be independent iff $P(A \cap B) = P(A) \cdot P(B)$, $A \in \mathcal{G}$, $B \in \mathcal{F}$

$A \in \mathcal{F}$, $\mathcal{F}_A = \{\phi, \Omega, A, A^c\}$ ← $\sigma$-algebra generated by the event $A$

**Proposition** $A, B$ independent iff $\mathcal{F}_A$, $\mathcal{F}_B$ independent

**Proposition** $\mathcal{A}$ any collection of subsets from $\Omega$. There exist the smallest $\sigma$-algebra which contains all elements form $\mathcal{A}$, $\sigma$-algebre ← $\sigma(\mathcal{A})$

**Definition 5** $\mathcal{A}$ be the colletion of all open sets in $R^n$, $\sigma(\mathcal{A})$ is the smallest $\sigma$-algebra contains all open sets in $R^n$, is called Borel $\sigma$-algebra on $\mathbb{R}^n$ or $\sigma$-algebra of Borel sets in $\mathbb{R}^2$ → $\mathcal{B}(\mathbb{R}^n)$

## (Lecture 4) A. Discrete RV / prob distribution

properties $\begin{cases} 0 \leq p(x) \leq 1, & x = 0, 1, \cdots \\ \sum_{x=0}^{\infty} p(x) = 1 \end{cases}$

A1. Binomial prob distribution (二项分布)

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \cdots, n$$

$X \sim B(n, p)$, $0 \leq p \leq 1$, $n > 0$ → pmf (prob mass function) 离散RV

$C_n^k = \dfrac{n!}{k! \times (n-k)!}$ $\boxed{\begin{matrix} 成功 p \\ 失败 1-p \end{matrix}}$

$E(X) = np$ $Var(X) = np(1-p)$

A2. Poisson distribution (泊松分布)

$$P(X=k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad \lambda > 0, k \in \{0, 1, \cdots\}$$

$X \sim Poisson(\lambda)$, $E(X) = Var(X) = \lambda$

$X \sim Poisson(\lambda_1)$, $Y \sim Poisson(\lambda_2)$

$X + Y \sim Poisson(\lambda_1 + \lambda_2)$

A3. Geometric prob distribution (几何分布)

$$P(X=k) = (1-p)^{k-1} \cdot p, \quad k = 1, 2, \cdots$$

$X \sim G(p)$, $E(X) = \dfrac{1}{p}$, $Var(X) = \dfrac{1-p}{p^2}$

## (Lecture 5) B. Continuous RV / prob distribution

( a real-valued function $f: R \to R$ satisfying $f(x) \geq 0$, $x \in R$, and $\int_R f(x) dx = 1$ )

$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(y) dy$

cdf (cumulative distribution function) 连续RV

pdf (prob density function)

B1. Normal (Gaussian) prob distribution (正态/高斯) $X \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$

$f(x) = \dfrac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in \mathbb{R}$

$E(X) = \mu$, $Var(X) = \sigma^2$

B2. Uniform distribution on $(a, b)$ (均匀分布)

$X \sim U([a, b])$. $E(X) = \dfrac{a+b}{2}$, $Var(X) = \dfrac{(b-a)^2}{12}$

$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x \notin [a, b] \end{cases}$

B3. Gamma prob distribution (gamma分布)

$X \sim \Gamma(\alpha, \beta)$, $f(x) = \begin{cases} \frac{1}{\Gamma(\alpha) \beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & x > 0 \\ 0, & x \leq 0 \end{cases}$

$\alpha > 0$, $\beta > 0$

Gamma function $\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx$

$\Gamma(1) = 1$, $\Gamma(a+1) = a\Gamma(a)$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

$\Gamma(a) = (a-1)!$

B4. Exponential distribution (指数分布)

$X \sim Exp(\lambda)$, $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$, $x \in [0, +\infty)$, $\lambda > 0$,

$F_X(x) = 1 - e^{-\lambda x}$, $x \in [0, +\infty)$, $E(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$

B5. $\chi^2$ prob distribution (卡方分布)

$X \sim N(0,1)$, $Y = X^2$, $F_Y(y) = P(Y \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_0^y \frac{1}{\sqrt{2\pi u}} e^{-u/2} du$

$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}$, $Y \sim \Gamma(\frac{1}{2}, 2) = \chi^2(1)$

$X \sim \chi^2(k)$. $f_X(x) = \frac{1}{2^{k/2} \Gamma(\frac{k}{2})} x^{k/2 - 1} e^{-x/2}$, $x \in [0, +\infty)$

B6. Rayleigh prob distribution (瑞利分布)

$X \sim Rayleigh(\sigma)$, $\sigma > 0$, $f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$, $x \in [0, +\infty)$

when $\sigma = 1$, $f(x) = x e^{-\frac{x^2}{2}}$ $f(x) \geq 0$, $\int f(x) dx = 1$

$X_1, X_2 \sim N(0,1)$ independent, $R = \sqrt{X_1^2 + X_2^2} \sim Rayleigh(1)$

## (Lecture 6) Random Vectors, prob distribution, Multidimensional cdf / pdf

$X = (X_1, X_2, \cdots, X_n): \Omega \to R^n$, where each $X_i: \Omega \to \mathbb{R}$ is random variable

$F_X(x_1, \cdots, x_n) = P(\omega: X_1(\omega) \leq x_1, \cdots, X_n(\omega) \leq x_n)$

$F_{X_i}(x_i) = P(X_i \leq x_i) := P(X_1 \in R, \cdots, X_i \leq x_i, \cdots, X_n \in R) = F_X(\infty, \cdots, x_i, \cdots, \infty)$

**Definition 1** $X_1, \cdots, X_n$ are independent iff $F_X(x_1, x_2, \cdots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$ or $f(x_1, \cdots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$

**Definition 2** $X = (X_1, \cdots, X_n)$ jointly continuous if there is a function $f: R^n \to R$ s.t.

1) $f(x) \geq 0$, all $x \in R^n$ 2) $\int_{R^n} f(x) dx = 1$, 3) $P(\omega: X(\omega) \in B) = \int_B f(x) dx$, all Borel set $B \in \mathcal{B}(R^n)$

$F_X(x_1, \cdots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(y_1, \cdots, y_n) dy_n \cdots dy_1$ joint pdf of vector $X$

**Definition 3** $X = (X_1, \cdots, X_n)$ → jointly continuous Gaussian distribution

$f_X(x) = \dfrac{1}{\sqrt{(2\pi)^n \det R}} \exp\left(-\frac{1}{2}(x-\mu)^T R^{-1} (x-\mu)\right)$, where $x = (x_1, \cdots, x_n) \in R^n$

vector $\mu \in R^n$, $R$, $n \times n$ positive definite matrix

**Simple example** $X_i \sim N(\mu_i, \sigma_i^2)$ independent RV.

$f_X(x) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) = \dfrac{1}{\sqrt{(2\pi)^n \det R}} \exp\left(-\frac{1}{2}(x-\mu)^T R^{-1} (x-\mu)\right)$

$\mu = (\mu_1, \cdots, \mu_n)$, $R$ diagonal matrix with element $\sigma_i^2$

## (Lecture 7) Transformation, Conditional densities

$$X = (X_1, \cdots, X_n) \underset{\varphi^{-1}}{\overset{\varphi}{\rightleftarrows}} Y = (Y_1, \cdots, Y_n)$$

$\varphi = (\varphi_1, \varphi_2, \cdots, \varphi_n)$ $\quad$ $Y = \varphi(X)$, $X = \varphi^{-1}(Y)$

Given joint pdf $f_X(x) \xrightarrow{?} f_Y(y)$ inverse transformation

$$P_Y(B) = P(Y = \varphi(X) \in B) = P(\varphi(X) \in B) = P(X \in \varphi^{-1}(B))$$

$$= \int_{\varphi^{-1}(B)} f_X(x) dx = \int_B f_X(\varphi^{-1}(y)) |\det D\varphi^{-1}(y)| dy$$

$x = \varphi^{-1}(y)$, $dx = |\det D\varphi^{-1}(y)| dy$

Jacobi matrix

$$\begin{pmatrix} \frac{\partial \varphi_1^{-1}(y_1, \cdots, y_n)}{\partial y_1} & \cdots & \frac{\partial \varphi_1^{-1}}{\partial y_n} \\ \vdots & & \\ \frac{\partial \varphi_n^{-1}}{\partial y_1} & \cdots & \frac{\partial \varphi_n^{-1}}{\partial y_n} \end{pmatrix}$$

$\Rightarrow f_Y(y) = f_X(\varphi^{-1}(y)) |\det D\varphi^{-1}(y)|$

**Example** $X = (X_1, X_2)$, $X_i \sim N(0,1)$ independent

polar coordinate, tansformation

$(X_1, X_2) \longrightarrow (R, \theta)$

$r = \varphi_1(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$, $\theta = \varphi_2(x_1, x_2) = \arctan \frac{x_2}{x_1}$

$x_1 = \varphi_1^{-1}(r, \theta) = r\cos\theta$, $x_2 = \varphi_2^{-1}(r, \theta) = r\sin\theta$

$|\det D\varphi^{-1}(r, \theta)| = \left\| \begin{matrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{matrix} \right\| = |r\cos^2\theta + r\sin^2\theta| = r$

$f_{X_1 X_2}(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{x_1^2 + x_2^2}{2}} \Longrightarrow f_{R,\theta}(r, \theta) = \frac{1}{2\pi} e^{-\frac{r^2}{2}} \cdot r$

$(r > 0, 0 \leq \theta < 2\pi)$

**Definition 1** conditional prob density

(1-dimensional pdf) → $f_{X_1|X_2=x_2}(x_1, x_2) = \dfrac{f_{X_1 X_2}(x_1, x_2)}{f_{X_2}(x_2)}$ $\quad \int f_{X_1|X_2=x_2}(x_1, x_2) dx_1 = 1$ ← 固定数

$P(X_1 \in B_1 | X_2 = x_2) = \int_B f_{X_1|X_2=x_2}(x_1, x_2) dx_1$

$g_1(x_2) = E(X_1 | X_2 = x_2) = \int x_1 f_{X_1|X_2=x_2}(x_1, x_2) dx_1$

**Example** $X, Y$ jointly continuous vector, independent, pdf: $f_X(x)$, $f_Y(y)$

define $Z = X + Y$. pdf of $Z$?

$F_Z(z) = P(X+Y \leq z) = \iint_{x+y \leq z} f_{XY}(x,y) dx dy \underset{(x=z-y)}{=} \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{z-y} f_X(x) dx \right) f_Y(y) dy = \int_{-\infty}^{+\infty} F_X(z-y) f_Y(y) dy \underset{\text{diff } z}{\Longrightarrow} f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y) dy$

## (Lecture 8) Expectation, Variance, properties

$X$ discrete RV. $EX = \sum_{i=1}^{\infty} x_i P(\omega: X(\omega) = x_i) = \sum_{i=1}^{\infty} x_i p(x_i)$

$X$ continuous RV. $EX = \int_R x f_X(x) dx$

**Example 1** a) $X \sim B(n, p)$, $EX = \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} = np$.

b) $X \sim Poisson(\lambda)$, $EX = \sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!}$ $(k = i-1)$

$= \lambda e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{e^\lambda} = \lambda e^{\lambda} e^{-\lambda} = \lambda$

c) $X \sim N(\mu, \sigma^2)$, $EX = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \underset{(z = \frac{x-\mu}{\sigma})}{=} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} (\mu + z\sigma) e^{-\frac{1}{2}z^2} dz$

$= \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz + \sigma \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} z e^{-\frac{1}{2}z^2} dz}_{\text{odd} \to 0} = \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz + 0\sigma = 1 \cdot \mu = \mu$

**Definition 1** $\phi: \mathbb{R} \to \mathbb{R}$ measurable map / function if $\phi^{-1}(B) \in \mathcal{B}(\mathbb{R})$ for all $B \in \mathcal{B}(\mathbb{R})$

$X$. RV. $\phi$ measurable function, $Y = \phi(X) \Rightarrow Y$ is again RV on $(\Omega, \mathcal{F}, P)$

**Properties of expectation**

1. $E(\phi(X)) = \int \phi(x) f_X(x) dx$ 2. $E(aX + bY) = aE(X) + bE(Y)$

3. $X, Y$ independent, $E(\phi(X)\varphi(Y)) = E(\phi(X)) \cdot E(\varphi(Y))$

**Definition 2** any $n = 1, 2, \cdots$, $\mu_n = E(X^n)$ called n-th moment of $X$

**Definition 3** $Var(X) = \sigma^2 = \sigma_X^2 = E[(X - \mu_X)^2]$

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