Lec 1

古典概型: $P(A)=\frac{N_A}{N}$ ← 组成事件A，← 可能数

局限: 需要所有情况等概率，且有限种情况.

频率概率: (Frequency probability)

$\lim\limits_{n\to\infty}\frac{n_A}{n}=p$, $P(A)=p$

(n trails)

Lec 2

样本空间 (sample space)

Ω : set of all possible outcome of an experiment

事件 (events) (用大写字母表示)

Any subset of the sample space

e.g. $A=\{9\}$ $E=\{x: 0\le x\le 5\}$

$A\cup B$, $A\cap B$, A^c, $\phi_i=\Omega^c$

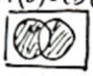
$A\cap B=\phi$, disjoint/mutually exclusive (互斥)

if $A\subset B$ and $B\subset A$, then $A=B$

$A\setminus B=A\cap B^c$

Symmetric difference: $A\triangle B=(A\setminus B)\cup(B\setminus A)$

($x\in A\cup B$ but $x\notin A\cap B$)



$\cap, \cup$ 满足交换律, 结合律, 分配律.

DeMorgan's Laws

$\left(\bigcup_{i=1}^{n}A_i\right)^c=\bigcap_{i=1}^{n}A_i^c$, $\left(\bigcap_{i=1}^{n}A_i\right)^c=\bigcup_{i=1}^{n}A_i^c$

$\boxed{\text{古典概型}}$: $P(\{i\})=\frac{1}{N}$ for each $1\le i\le N$

For any $E\in\mathcal{F}$, $P(E)=\frac{|E|}{N}$

~~排列~~ 组合数

$\binom{n}{m}=C_n^m=\frac{n!}{m!(n-m)!}$

若A, B互斥: $|A\cup B|=|A|+|B|$

$P(A\cup B)=P(A)+P(B)$

若 $A_1,\dots A_n$ 互斥, $P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)$

$P(\Omega)=1$, Ω 是必然事件.

For any $A\subset\Omega$, $p(A)\ge 0$

e.g. $P(A^c)=1-P(A)$

$P(A\cup B)=P(A)+P(B)-P(A\cap B)\le P(A)+P(B)$

$P(A\cap B)+P(A\setminus B)=P(A)$

if $B\subset A$, $P(A\setminus B)=P(A)-P(B)\ge 0$

容斥原理: $P(E_1\cup E_2\cup\dots\cup E_n)$

$=\sum_{i=1}^{n}P(E_i)+\dots+(-1)^{r+1}\sum_{1\le i_1<i_2<\dots<i_r\le n}P(E_{i_1}\cap E_{i_2}\cap\dots\cap E_{i_r})$

$+\dots+(-1)^{n+1}P(E_1\cap\dots\cap E_n)$

匹配问题: n个人n顶帽子, 没有人拿到自己帽子的概率

E_i: 第i个人拿到自己的. 有人拿到自己帽子的

$P(E_1\cup E_2\cup\dots\cup E_n)=\dots$

$P(E_{i_1}\cap E_{i_2}\cdots\cap E_{i_r})=\frac{(n-r)!}{N!}$

Ans: $\sum_{i=0}^{N}\frac{(-1)^i}{i!}$, $N\to\infty$, $P\to\frac{1}{e}$ $\left(e^x=\sum_{i=0}^{\infty}\frac{x^i}{i!}\right)$

$\boxed{\text{Probability on a general set}}$:

$\Omega=\{w_j, j\in J\}$ (可数), $\mathcal{F}=2^{\Omega}$. Choose any sequence of numbers $\{p_j, j\in J\}$ satisfying that

$\forall j\in J$: $p_j\ge 0$; $\sum_{j=1}p_j=1$

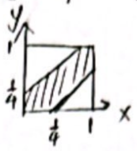
Def a set function $P:\mathcal{F}\to[0,1]$ as $\forall E\in\mathcal{F}$: $P(E)=\sum_{w_j\in E}p_j$

Then $(\Omega,\mathcal{F},P)$ is a probability space.

probability measure individual's degree of belief

几何概率

$P(A)=\frac{\text{Geometric measure of }A}{\text{Geometric measure of }\Omega}$



两个人在某时段内到达…, 两人有可能在0~1小时, 第1次到达后等待…, 问: 相遇概率.

$\Omega=[0,1]\times[0,1]$

$E=\{(x,y):|x-y|\le\frac{1}{4}, 0\le x,y\le 1\}$

Axiom of probability:

(i) $P(E)\ge 0$ for all $E\in\mathcal{F}$ (Non-negativity)

(ii) $P(\Omega)=1$ (Normalization)

(iii) mutually exclusive $E_1,E_2\cdots$ (σ-additivity)

$P\left(\bigcup_{i=1}^{\infty}E_i\right)=\sum_{i=1}^{\infty}P(E_i)$

(prop 具体看上栏).

补充:

Boole's inequality: for $E_1,E_2\cdots\in\mathcal{F}$ (不一定互斥)

$\sum_{i=1}^{n}P(E_i)-\sum_{i<j}P(E_iE_j)\le P\left(\bigcup_{i=1}^{n}E_i\right)\le\sum_{i=1}^{n}P(E_i)$

补充:

① $P(A\cap B)\ge P(A)+P(B)-1$

$\Leftrightarrow 1\ge P(A)+P(B)-P(A\cap B)=P(A\cup B)$

② $P\left(\bigcap_{i=1}^{n}A_i\right)\ge\sum_{i=1}^{n}P(A_i)-(n-1)$

$\Leftrightarrow n-\sum_{i=1}^{n}P(A_i)\ge 1-P\left(\bigcap_{i=1}^{n}A_i\right)=P\left(\left(\bigcap_{i=1}^{n}A_i\right)^c\right)$

$\sum_{i=1}^{n}(1-P(A_i))=\sum_{i=1}^{n}P(A_i^c)$ $=P\left(\bigcup_{i=1}^{n}A_i^c\right)$

③ $A=A_1\cup\dots\cup A_n$, $S_k=\sum_{1\le i_1<\dots<i_k\le n}P(A_{i_1}\cap A_{i_2}\cap\dots\cap A_{i_k})$

$P(A)\le\sum_{k=1}^{m}(-1)^{k-1}S_k$, for odd m

$P(A)\ge\sum_{k=1}^{m}(-1)^{k-1}S_k$, for even m

Lec 3

条件概率: $P(E|F)=\frac{P(E\cap F)}{P(F)}$

$\tilde{P}(\cdot):\mathcal{F}\to[0,1]$ as $\tilde{P}(E)=P(E|F)$

Then $(\Omega,\mathcal{F},\tilde{P})$ is also a probability space

$\tilde{P}(\cdot|F)$ 满足所有 general property 的性质

Multiplication rule:

$P(E\cap F)=P(E|F)P(F)$

$P(E_1\cap E_2\cap\dots\cap E_n)=P(E_1)P(E_2|E_1)P(E_3|E_1\cap E_2)\cdots P(E_n|E_1\cap E_2\cap\dots\cap E_{n-1})$

全概率公式:

$E=(E\cap F)\cup(E\cap F^c)$

$P(E)=P(E\cap F)+P(E\cap F^c)$

$=P(E|F)P(F)+P(E|F^c)P(F^c)$

~~若E_1~~, 若 $F_1,F_2,\cdots$ 是互斥且是Ω的partition,

$\Omega=\bigcup_{i=1}^{\infty}F_i$, 则

$P(E)=\sum_{i=1}^{\infty}P(F_i)P(E|F_i)$

Bayes' formula

Let $E_1,E_2\cdots$ be a partition of Ω, for $P(F)>0$

$P(E_i|F)=\frac{P(E_i)P(F|E_i)}{P(F)}$

$=\frac{P(F|E_i)P(E_i)}{\sum_{j=1}^{\infty}P(F|E_j)P(E_j)}$

独立事件: $P(A\cap B)=P(A)P(B)$, 则A, B相互独立

$P(B|A)=P(B)$

prop: 若E, F独立, 则E和F^c也独立

~~相互独立~~

① $P(E\cap F\cap G)=P(E)P(F)P(G)$ (mutually/jointly)

② $P(E\cap F)=P(E)P(F)$ 相互独立: ①②③④

③ $P(E\cap G)=P(E)P(G)$ 两两独立: ②③④

④ $P(F\cap G)=P(F)P(G)$ pair-wise

General independence:

$P(E_1'E_2'\cdots E_r')=P(E_1')P(E_2')\cdots P(E_r')$

$r\le n$, (E_r')是(E_n)的子集.

补充: 若A, B, C相互独立, 则$A\cap B$和C, $A\cup B$和C都相互独立

if P is σ-additive, if $A_n\in\mathcal{F}$, $A_n\downarrow\phi$, then $P(A_n)\downarrow 0$

Lec 4

Random variable: $X:\Omega\to\mathbb{R}$

离散随机变量, S is finite or countable infinite

probability mass function (pmf):

$p(a)=P\{X=a\}$ for $a\in S$

prop: (1) $p(s_i)\ge 0$ (2) $\sum p(s_i)=1$

可绘成图、表

(Cumulative) distribution function (cdf):

$F(x)=P\{X\le x\}$

prop.

(i) non-decreasing (ii) right continuous

(iii) $\max=\lim\limits_{x\to\infty}F(x)=1$ (iv) $\min=\lim\limits_{x\to-\infty}F(x)=0$

if $a<b$, then $P\{a<X\le b\}=F(b)-F(a)$

$P\{X<b\}=F(b-)=\lim\limits_{n\to\infty}F(b-\frac{1}{n})$

$P\{X\ge a\}=1-F(a-)$ $P\{X=a\}=F(a)-F(a-)$

function of r.v.

$P\{g(X)=y\}=P\{X=g^{-1}(y)\}$ if g one-to-one

$P\{g(X)=y\}=\sum_{x:g(x)=y}P\{X=x\}$

期望:

$E[X]=\sum_{j\in J}s_jp(s_j)=\sum_{x\in S}xp(x)$

$\sum_{x\in S}|x|p(x)<\infty$, then exp is well-defined.

$E[g(X)]=\sum_{x\in S}g(x)p(x)$

if $f(X)\le g(X)$, then $E[f(X)]\le E[g(X)]$

$E[af(X)+bg(X)]=aE[f(X)]+bE[g(X)]$

$E[aX+b]=aE[X]+b$

$E\left[\sum_{i=1}^{n}X_i\right]=\sum_{i=1}^{n}E[X_i]$

$E[g(X)]$ is larger if g is convex ($E[g(x)]\ge g(E[X])$)

convex: $f(\lambda x+(1-\lambda)y)\le\lambda f(x)+(1-\lambda)f(y)$

e.g. $|x|$, x^2, $|x|^p$ for $p\ge 1$, $e^{\lambda x}$, $\max\{x,a\}$

方差:

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

$Var(aX+b)=a^2Var(X)$

Loss function: $l(s)=(x-s)^2$

$L(s)=E[(X-s)^2]$ $L(\mu)=Var(X)$

标准差: $SD(X)=\sqrt{Var(X)}$

Mean absolute deviation: $MAD(X)=E[|X-\mu|]$

$MAD(X)\le SD(X)$

中位数：取中间位置，若有两个则取平均.

$F(m) \geq \frac{1}{2}$ $F(m-) \leq \frac{1}{2}$ then $med(X)=m$

Bernoulli trail: outcome success or failure.

Bernoulli distribution:

$X \sim Bernoulli(p)$

$E[X]=p$, $Var(X)=p(1-p)$

Binomial distribution: n重Bernoulli trails.

$X \sim B(n,p)$

pmf: $p(k)=\binom{n}{k}p^k(1-p)^{n-k}$ for $k=0,1,\cdots,n$

$E[X]=np$, $Var(X)=np(1-p)$

$p(k)$在 $k=\lfloor(n+1)p\rfloor$ 处达到最大(比值法)

Geometric distribution:

$X \sim Geometric(p)$ ~~[illegible]~~

pmf: $p(k)=(1-p)^{k-1}p$, $k=1,2\cdots$

cdf: $P\{X \geq k\}=(1-p)^{k-1}$, $k=1,2\cdots$

$E[X]=\frac{1}{p}$ $Var(X)=\frac{1-p}{p^2}$

prop memoryless

$P\{X>s+t \mid X>s\}=P\{X>t\}$

$\frac{(1-p)^{s+t}}{(1-p)^s}=(1-p)^t$

Poisson distribution: p很小, n很大

$X \sim Poisson(\lambda)$

pmf: $p(k)=e^{-\lambda}\frac{\lambda^k}{k!}$ for $k=0,1,2\cdots$

$E[X]=\lambda$, $Var(X)=\lambda$

$P\{X=k\}=\binom{n}{k}p^k(1-p)^{n-k}=\frac{n!}{k!(n-k)!}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k}$

for large n and moderate λ. $=\frac{(n)_k}{n^k}\frac{\lambda^k}{k!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}$

$(1-\frac{\lambda}{n})^n \approx e^{-\lambda}$, $\frac{(n)_k}{n^k}\approx 1$, $(1-\frac{\lambda}{n})^k \approx 1$

$\therefore P\{X=k\}\approx e^{-\lambda}\frac{\lambda^k}{k!}$

补充:

hypergeometric distribution: N个球, m个白, 取n个

$p(i)=\frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$ $i=0,\cdots,m$

$p=\frac{m}{N}$, $E[X]=np$, $Var(X)=\frac{N-n}{N-1}np(1-p)$

negative binomial distribution:

$p(i)=\binom{i-1}{r-1}p^r(1-p)^{i-r}$ $i \geq r$

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

Lec 5.

连续随机变量

$P\{X\in B\}=\int_B f(x)dx$

pdf: probability density function

prop (i) $f(x)\geq 0$ for all $x\in\mathbb{R}$

(ii) $\int_{-\infty}^{\infty} f(x)dx=\int_S f(x)dx=1$

cdf: $F(x)=P\{X\leq x\}=\int_{-\infty}^{x} f(t)dt$

(i) for $x<y$, $P\{x<X\leq y\}=F(y)-F(x)=\int_x^y f(u)du$

(ii) F is continuous

(iii) $P\{X=x\}=0$

$\frac{d}{dx}F(x)=f(x)$

4 $Y=g(X)$

$F_Y(y)=P\{Y\leq y\}=P\{g(X)\leq y\}$ $=P\{X\in g^{-1}(-\infty,y]\}$

$^*P\{X\leq g^{-1}(y)\}$ (g递增)

Example: $Y=X^2$

$F_Y(y)=P\{Y\leq y\}=P\{X^2\leq y\}$

$=P\{-\sqrt{y}\leq X\leq\sqrt{y}\}=F_X(\sqrt{y})-F_X(-\sqrt{y})$

$f_Y(y)=F'_Y(y)=\frac{1}{2\sqrt{y}}\{f_Y(\sqrt{y})-f_Y(-\sqrt{y})\}$

$E[X]=\int_{-\infty}^{\infty}xf(x)dx$ ($\int_{-\infty}^{\infty}|x|f(x)dx<\infty$)

$E[g(X)]=\int_S g(x)f(x)dx$

Lemma: $E[Y]=\int_0^{\infty}P\{Y>y\}dy$

proof: $=\int_0^{\infty}\int_y^{\infty}f_Y(x)dxdy=\int_0^{\infty}(\int_0^x dy)f_Y(x)dx$

$=\int_0^{\infty}xf_Y(x)dx=E[Y]$

$E[g(X)]=\int_0^{\infty}P\{g(X)>y\}dy$

$=\int_0^{\infty}\int_{x:g(x)>y}f(x)dxdy=\int_{x:g(x)>0}\int_0^{g(x)}dy f(x)dx$

$=\int_{x:g(x)>0}g(x)f(x)dx$

$E[ag(X)+bh(X)]=aE[g(X)]+bE[h(X)]$

$E[aX+bY]=aE[X]+bE[Y]$

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

$Var(aX+b)=a^2Var(X)$

Uniform distribution. $X\sim Uniform(a,b)$

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

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

Normal distribution: $X\sim N(\mu,\sigma^2)$

pdf: $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty<x<\infty$

prop. $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-x^2/2}dx=1$ ~~take $y=\frac{x-\mu}{\sigma}$~~

$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-y^2/2}dy=1$

$aX+b\sim N(a\mu+b,a^2\sigma^2)$

$\frac{X-\mu}{\sigma}\sim N(0,1)$

则 $E[X]=\mu$, $Var[X]=\sigma^2$

$X\sim N(0,1)$: standard normal random variable.

cdf: $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt$ (可查表)

$\Phi(-x)=1-\Phi(x)$ $P\{X\leq x\}=\Phi(x)$

if $X\sim N(\mu,\sigma^2)$, then

$P\{X\leq a\}=P\{\frac{X-\mu}{\sigma}\leq\frac{a-\mu}{\sigma}\}=\Phi(\frac{a-\mu}{\sigma})$

Exponential distribution: $X\sim Exp(\lambda)$

pdf: $f(x)=\begin{cases}\lambda e^{-\lambda x} & x\geq 0\\ 0 & x<0\end{cases}$

cdf: $F(x)=\begin{cases}1-e^{-\lambda x}, & x\geq 0\\ 0 & x<0\end{cases}$

$E[X]=\frac{1}{\lambda}$, $Var(X)=\frac{1}{\lambda^2}$

prop. memoryless

$P\{X>s+t \mid X>t\}=P\{X>s\}$

常用于等到某一事件发生的时间.

Γ-distribution: (gamma) (α,λ)

pdf: $f(x)=\begin{cases}\frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}=\frac{x^{\alpha}}{\Gamma(\alpha)}e^{-\lambda x}x^{\alpha-1}, & x\geq 0\\ 0, & x<0\end{cases}$

prop. $\Gamma(\alpha)=\int_0^{\infty}e^{-y}y^{\alpha-1}dy$ (gamma function)

$\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$ if $\alpha=n$, $\Gamma(n)=(n-1)!$

$E[X]=\frac{\alpha}{\lambda}$, $Var(X)=\frac{\alpha}{\lambda^2}$

Cauchy distribution.

pdf: $f(x)=\frac{1}{\pi}\frac{1}{1+x^2}$, $-\infty<x<\infty$

Beta distribution:

pdf: $f(x)=\begin{cases}\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}, & 0<x<1\\ 0, & \text{otherwise.}\end{cases}$

$B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}dx=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

$E[X]=\frac{a}{a+b}$ $Var(X)=\frac{ab}{(a+b)^2(a+b+1)}$

对于单调函数$g(x)$

$Y=g(X)$, then

$f_Y(y)=\begin{cases}f_X[g^{-1}(y)]\left|\frac{d}{dy}g^{-1}(y)\right| & \text{if } y=g(x) \text{ for some } x\\ 0 & \text{if } y\neq g(x) \text{ for all } x\end{cases}$

Th. 若F严格递增且连续, 则

$Y:=F(X)\sim U(0,1)$

Inverse function $F^{-1}(\cdot)$

$F^{-1}(u)=\inf\{x: F(x)\geq u\}$

Th. 若$U\sim U(0,1)$, Then $X=F^{-1}(U)$ has the distribution F