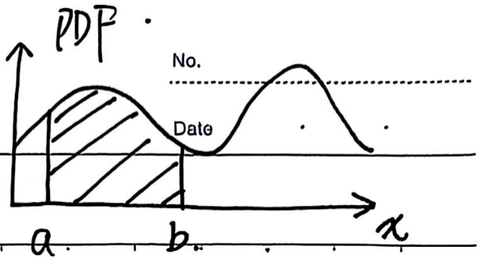


第五章总结：连续型随机变量

一. Probability Density Functions

之前一直是离散型随机变量，它的CDF（累积分布函数）在取值点处有跳跃。那么连续呢？它的曲线将较光滑

· Def：若r.v.累积分布函数可微，则可称r.v.服从一个连续分布。此为允许存在这样的端点，在这些点CDF连续而不可微。

这个CDF的导数有大用处！它被称为概率密度函数PDF！

· Def：连续型r.v. X，CDF为F，则其导数为PDF

PDF是类似于PMF的东西吗？类似，但有大区别！连续r.v.中$P(X=x_0)=0$，因为X的CDF没有跳跃点，因此PDF上点与某事件的概率无关；甚至PDF在一些x取值上可以大于1。PDF的意义是为了计算区间内概率。

· Prop：连续r.v. X的PDF为f，则X的CDF为$F(x)=\int_{-\infty}^{x} f(t)dt$

类似于通过把PMF在所有$\leq x$点处值求和以得CDF在x处取值

· Theorem：求x落在(a,b)间概率：

$$P(a<x<b)=P(a\leq x<b)=P(a<x\leq b)=P(a\leq x\leq b)=\int_a^b f(x)dx$$，见fig1

△：$P(x=a)=P(X=b)=0$

更特殊地：X, Y为continuous r.v.，若X, Y独立，则$P(X=Y)=0$.

Proof：$P(X=Y=\lambda)=P(X=\lambda)\cdot P(Y=\lambda)=0\cdot 0=0$

⚠但X, Y不独立，则是另一码事！此条不成立。

那么一个PDF是否valid？它必须满足：

① 非负：$f(x)>0$　② 积为1：$\int_{-\infty}^{+\infty} f(x)dx=1$

例如：Logistic Distribution，以之为例：

CDF：$F(x)=\frac{e^x}{1+e^x}$，$x\in R$

PDF：$f(x)=F'(x)=\frac{e^x}{(1+e^x)^2}$，$x\in R$

$$\int_{-\infty}^{+\infty} f(x)dx = F(x)\Big|_{-\infty}^{+\infty} = 1-0=1$$

No.

Date

又例：Rayleigh Distribution：

CDF： $F(x) = 1 - e^{-\frac{x^2}{2}}$, $x>0$

PDF： $f(x) = F'(x) = xe^{-\frac{1}{2}x^2}$, $x>0$; $\int_{-\infty}^{+\infty} f(x)dx = F(x)\Big|_0^{+\infty} = 1$

那么类似于Discrete r.v.，Continous r.v.期望如何计算？

Def：Cont. r.v. X，PDF为f，则期望为：

$$E(X) = \int_{-\infty}^{+\infty} x f(x)dx$$

Theorem：(无意识统计)，$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x)dx$

Theorem：Expectation via Survival Function：Cont. r.v X，非负，F为CDF，则令 $G(x) = 1-F(x) = P(X>x)$，则G称为Survival Function，有：

$$E(X) = \int_0^{\infty} G(x)dx$$

→ $P(X>x)$ 可以很好帮助理解 $G(x)$ 意义！

之前提及了端点处P为0的结论，它能引申许多推论：

Theorem：$X_1, \cdots, X_n$ be i.i.d from Cont. Distribution. 则 $P(X_{a_1} < \cdots < X_{a_n}) = 1/n!$ for any permutation $a_1, \cdots, a_n$ of $1, \cdots, n$.

二. Uniform Distribution

• Def：Cont. r.v U在(a,b)区间中服从均匀分布若：

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

记作：$U \sim Unif(a,b)$

更特别地：$U \sim Unif(0,1)$，$P(U \le x) = \begin{cases} 0, & x \le 0 \\ x, & 0<x<1 \\ 1, & x \ge 1. \end{cases}$

例题：$X_1, \cdots, X_n$ are i.i.d $Unif(0,1)$，

$Y = \min\{X_1, \cdots, X_n\}$，求 $E(Y)$.

正向求很难！可试Survival！ $G(Y) = P(Y \ge y) = P(X_1 \ge y) \cdots P(X_n \ge y) = (1-y)^n$

$\therefore E(Y) = \int_0^1 (1-y)^n dy = \int_0^1 y^n dy = \frac{1}{n+1}$

三. Basic Monte Carlo Simulation ⟷ Universality of Uniform

利用 Unif(0,1) r.v.，我们能构建任何一个Cont. Dist. r.v.
同样，用 Cont. Dist. r.v.，我们也能构 Unif(0,1) r.v.

· Theorem：F是 CDF，F严格单增以确保 F^{-1} 存在，则有：

$$\begin{cases} U\sim \text{Unif}(0,1),\ X=F^{-1}(U)，\text{则 } X \text{ 是以 } F \text{ 为CDF的 r.v.} \\ \text{让 } X \text{ 是以 } F \text{ 为CDF的 r.v.，则 } F(X)\sim \text{Unif}(0,1) \end{cases}$$

Proof：① $X=F^{-1}(U)$. 则 $P(X\le x)=P(F^{-1}(U)\le x)$
$= P(U\le F(x)) \overset{\text{Unif性质}}{=} F(x)$

② $P(Y\le y)=P(F(X)\le y)=P(X\le F^{-1}(y))=F(F^{-1}(y))=y$
故 $Y\sim \text{Unif}(0,1)$.

Eg：在 Logistic中的应用：$F(x)=\frac{e^x}{1+e^x}$，$x\in\mathbb{R}$
$U\sim \text{Unif}(0,1)$. $F^{-1}(u)=\ln(\frac{u}{1-u})$，U换u：
$F^{-1}(U)=\ln(\frac{U}{1-U})$，则：$\ln(\frac{U}{1-U})\sim$ logistic
Proof：$P(\ln(\frac{U}{1-U})\le x)=P(\frac{U}{1-U}\le e^x)=P(U\le\frac{e^x}{1+e^x}) \overset{\text{Unif性质}}{=} \frac{e^x}{1+e^x}$

同理在 Rayleigh中应用：$F(x)=1-e^{-\frac{x^2}{2}}$
$F^{-1}(u)=\sqrt{-2\ln(1-u)}$，则 $\sqrt{-2\ln(1-U)}\sim$ Rayleigh
Proof：$P(\sqrt{-2\ln(1-U)}\le x)=P(U\le 1-e^{-\frac{x^2}{2}})=1-e^{-\frac{x^2}{2}}$

四. 指数分布 Exponential Distribution

· Def：若连续 r.v. X 的 PDF为：

$$f(x)=\lambda e^{-\lambda x},\quad x>0$$

则称 X 满足参数为λ的指数分布，$X\sim \text{Expo}(\lambda)$
且它的 CDF为：$F(x)=1-e^{-\lambda x}$，$x>0$

· 方差与期望：$Y\sim \text{Expo}(\lambda)$，则：
$E(Y)=\frac{1}{\lambda}$　　$Var(Y)=\frac{1}{\lambda^2}$

No.

Date

此外，指数分布有一个非常特殊的性质：无记忆性

· Def：若r.v. X 对 $\forall s,t>0$，满足：

$$P(X\geq s+t \mid X\geq s)=P(X\geq t)$$

则称该分布有无记忆性

那下面说明：$X\sim Expo(\lambda)$ 为何有"无记忆性"：

CDF：$F(x)=1-e^{-\lambda x}$，则 $P(X\geq t)=1-(1-e^{-\lambda(t)})=e^{-\lambda t}$

$$\therefore P(X\geq s+t \mid X\geq s)=\frac{P(X\geq s+t)}{P(X\geq s)}=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}=e^{-\lambda t}=P(X\geq t)$$

△：Theorem：若 X 是无记忆的正连续型随机变量，则 $X\sim Expo(\lambda)$

Theorem：Minimum of Independent Expos：

$X_1,\cdots,X_n$ 独立，$X_j\sim Expo(\lambda_j)$．令 $L=\min\{X_1,\cdots,X_n\}$．

则：$L\sim Expo(\lambda_1+\cdots+\lambda_n)$．

Proof：$P(X_j\geq t)=e^{-\lambda t}$，则 $P(L\geq t)=P(X_1\geq t,\cdots,X_n\geq t)$

$=P(X_1\geq t)\cdots P(X_n\geq t)=e^{-(\lambda_1+\lambda_2+\cdots)t}$

则 L 的CDF：$1-e^{-(\lambda_1+\cdots+\lambda_n)x}$，故 $L\sim Expo(\lambda_1+\cdots+\lambda_n)$．

另外：补充 Failure Rate Function 概念

X 是 Cont. r.v.，PDF为 $f(t)$，CDF为 $F(t)=P(X\leq t)$，

则 FRF $r(t)$ 是：

(失败率函数)

$$r(t)=\frac{f(t)}{1-F(t)}$$

它指某部件工作到某一时刻尚未失效，在该时刻后单位时间内失效的概率

No.

Date

五. 正态分布

· Def: r.v. Z 被认为有标准正态分布 if PDF φ is:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

我们记作 $Z \sim N(0,1)$，因为 $\varphi(z)$ [平均值为0，方差为1] 后续证明

CDF Φ: $\phi(z) = \int_{-\infty}^{z} \varphi(t) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$

· 由 PDF. CDF 可知以下性质:

$$\varphi(z) = \varphi(-z) \qquad \phi(z) + \phi(-z) = 1$$

if $Z \sim N(0,1)$ then $-Z \sim N(0,1)$

· Validity: $\left(\int_{-\infty}^{+\infty} e^{-z^2/2} dz\right)^2 = \int_{-\infty}^{+\infty} e^{-x^2/2} dx \int_{-\infty}^{+\infty} e^{-y^2/2} dy$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}} dx dy = \int_0^{2\pi}\int_0^{+\infty} e^{-r^2/2} r dr d\theta$$

$$= \int_0^{2\pi}\left(\int_0^{+\infty} e^{-u} du\right) d\theta = 2\pi \quad \therefore \lim_{z\to\infty} \phi(z) = 1$$

· $E(Z) = 0$，显然，因为 $\varphi(z) = \varphi(-z)$ (对称性)

$$Var(Z) = E(Z^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 e^{-z^2/2} dz = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} z^2 e^{-z^2/2} dz$$

$u = z$, $dv = z e^{-z^2/2} dz$，则 $du = dz$, $v = -e^{-z^2/2}$，则:

$$\frac{2}{\sqrt{2\pi}} \cdot \left(-z e^{-z^2/2}\right)\Big|_0^{+\infty} + \int_0^{+\infty} e^{-z^2/2} dz$$

$$= \frac{2}{\sqrt{2\pi}}\left(0 + \frac{\sqrt{2\pi}}{2}\right) = 1.$$

· Def: If $Z \sim N(0,1)$, then:

$$X = \mu + \sigma Z$$

where μ is mean and σ^2 is variance

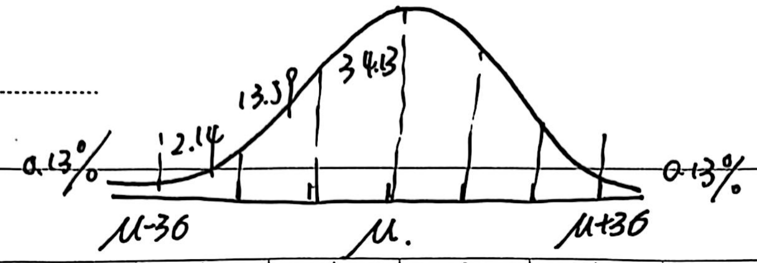
CDF: $F(x) = \phi\left(\frac{x-\mu}{\sigma}\right)$

PDF: $f(x) = \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$

KOKUYO

No.

Date



这个钟形曲线如上图，可以观察到 置信区间：

$P(|X-\mu|<\sigma)\approx 0.68$ $\quad$ $P(|X-\mu|<2\sigma)\approx 0.95$

$P(|X-\mu|<3\sigma)\approx 0.997$

六、中心极限定理

· Def: Let $X_1, \cdots, X_n$ be i.i.d r.v with finite μ & σ^2, then the sample mean $\overline{X_n}$ is defined as follows:

$$\overline{X_n} = \frac{1}{n}\sum_{j=1}^{n} X_j ,$$

then $E(\overline{X_n}) = \mu$, $Var(\overline{X_n}) = \frac{\sigma^2}{n}$

(notice that: $n\to\infty$, $Var(\overline{X_n})\to 0$)

· 中心极限定理： Central Limit Theoram

$n\to\infty$, $\sqrt{n}\left(\frac{\overline{X_n}-\mu}{\sigma}\right) \longrightarrow N(0,1)$

i.e., for large n , $\overline{X_n}$ 分布近似于 $N(\mu, \sigma^2/n)$

i.e., $n\overline{X_n} = \sum_{j=1}^{n} X_j$ 近似于 $N(n\mu, n\sigma^2)$

这是一个很有意思的分布定理。同一分布（同 μ 同 σ^2）的 r.v. 取平均，在 $n\to\infty$ 下，这个 r.v. 竟然趋于 $N(\mu, \sigma^2/n)$！

· $Pois(n)\to N$: $\quad Y\sim Pois(n)$ ，可认为是 $\sum_{i=1}^{n} Pois(1)$,

则 $n\to\infty$, $Y\to N(n, n)$.

$Binom\to N$: $\quad Y\sim Bin(n,p)$. 可认为是 $\sum_{i=1}^{n} Bin(1,p)$ ($Bern(p)$).

则 $n\to\infty$, $Y\to N(np, np(1-p))$.

但是，Y有离散性啊！Normal 分布是连续域的！

因此有 连续性 校正：

No.
Date

Continuity Correction: De Moivre-Laplace Approximation

$P(Y=k)=P(k-\frac{1}{2}<Y<k+\frac{1}{2})\approx\phi\left(\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)-\phi\left(\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)$

这个近似很work：$n\leq 50$下，$p\sim 1/2$时，近似非常好！

这么NB的Theorem怎么证？首先要了解一个东西：MGF.

七. MGF - Moment Generating Function 矩母函数.

· Def: MGF of a r.v. X is:

$M(t)=E(e^{tX})$ as a function of t

Eg: Bern: $M(t)=(1-p)+pe^t$

Unif: $M(t)=\int_a^b\frac{1}{b-a}e^{tu}du=\frac{e^{tb}-e^{ta}}{t(b-a)}$.

我们可以认为：MGF encodes the distribution of an r.v.

Theorem: Given the MGF of X, we can get the n^{th} moment of X by evaluating the n^{th} derivative of MGF at 0:

$E(X^n)=M^{(n)}(0)$.

这一应用，在这前面~~tossing~~ dies中已有使用

Theorem: r.v.的MGF决定了它的分布

If X, Y独立, then: $M_{X+Y}(t)=M_X(t)M_Y(t)$.

If X有MGF: $M(t)$, 则 $a+bX$: $E(e^{t(a+bX)})=e^{at}E(e^{btX})$

$=e^{at}M(bt)$.

Eg: Poisson MGF: $X\sim Pois(\lambda)$, $E[e^{tX}]=\sum_{k=0}^{\infty}e^{tk}\cdot\frac{e^{-\lambda}\lambda^k}{k!}$

$=e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^t)^k}{k!}=e^{\lambda(e^t-1)}$

☆ For $X\sim N(\mu,\sigma)$ $M_X(t)=e^{\mu t+\frac{1}{2}\sigma^2t^2}$

KOKUYO

No.

Date

这种MGF与 Generating function $E(t^x)=\sum_{k=0}^{\infty} p_k t^k$. 不同

，后者称为 PGF (Probability).

△: PGF处理 discrete，MGF处理 general r.v.

回到 CLT (极限中心) 证明:

$$\sqrt{n}\left(\frac{\bar{X}_n-\mu}{\sigma}\right)\to N(0,1).$$

Proof: 令 $M(t)=E(e^{tX_j})$，$M(0)=1$, $M'(0)=\mu=0$, $M''=\sigma^2=1$

则 $E(e^{t(X_1+\cdots+X_n)/\sqrt{n}})=E(e^{tX_1/\sqrt{n}})\cdots=\left(M\left(\frac{t}{\sqrt{n}}\right)\right)^n$

(因为欲证: $\sqrt{n}\bar{X}_n\to N(0,1)$, i.e., $M_{\sqrt{n}\bar{X}_n}(t)\to M_{N(0,1)}(t)$).

取对数后取极限:

$$\lim_{n\to\infty} n\ln M\left(\frac{t}{\sqrt{n}}\right)\overset{y=1/\sqrt{n}}{=}\lim_{y\to 0}\frac{\ln(M(yt))}{y^2}=\lim_{y\to 0}\frac{tM'(yt)}{2yM(yt)}$$

$$=\frac{t}{2}\lim_{y\to 0}\frac{M'(yt)}{y}=\frac{t^2}{2}\lim_{y\to 0}M''(yt)=\frac{t^2}{2}$$

$\therefore n\to\infty$, $\left(M\left(\frac{t}{\sqrt{n}}\right)\right)^n\to e^{t^2/2}$，这里正是正态分布MGF

因为: $M_Y(t)=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}e^{ty}dy=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{y^2}{2}+ty}dy$

$$=e^{\frac{t^2}{2}}\frac{1}{\sqrt{2\pi}}\underbrace{\int_{-\infty}^{\infty}e^{-\frac{(y-t)^2}{2}}dy}_{=\sqrt{2\pi}\ (\text{前面已证})}=e^{t^2/2}$$