Lecture 5: Continuous Random Variables

Wen Dingzhu

School of Information Science and Technology (SIST)
ShanghaiTech University

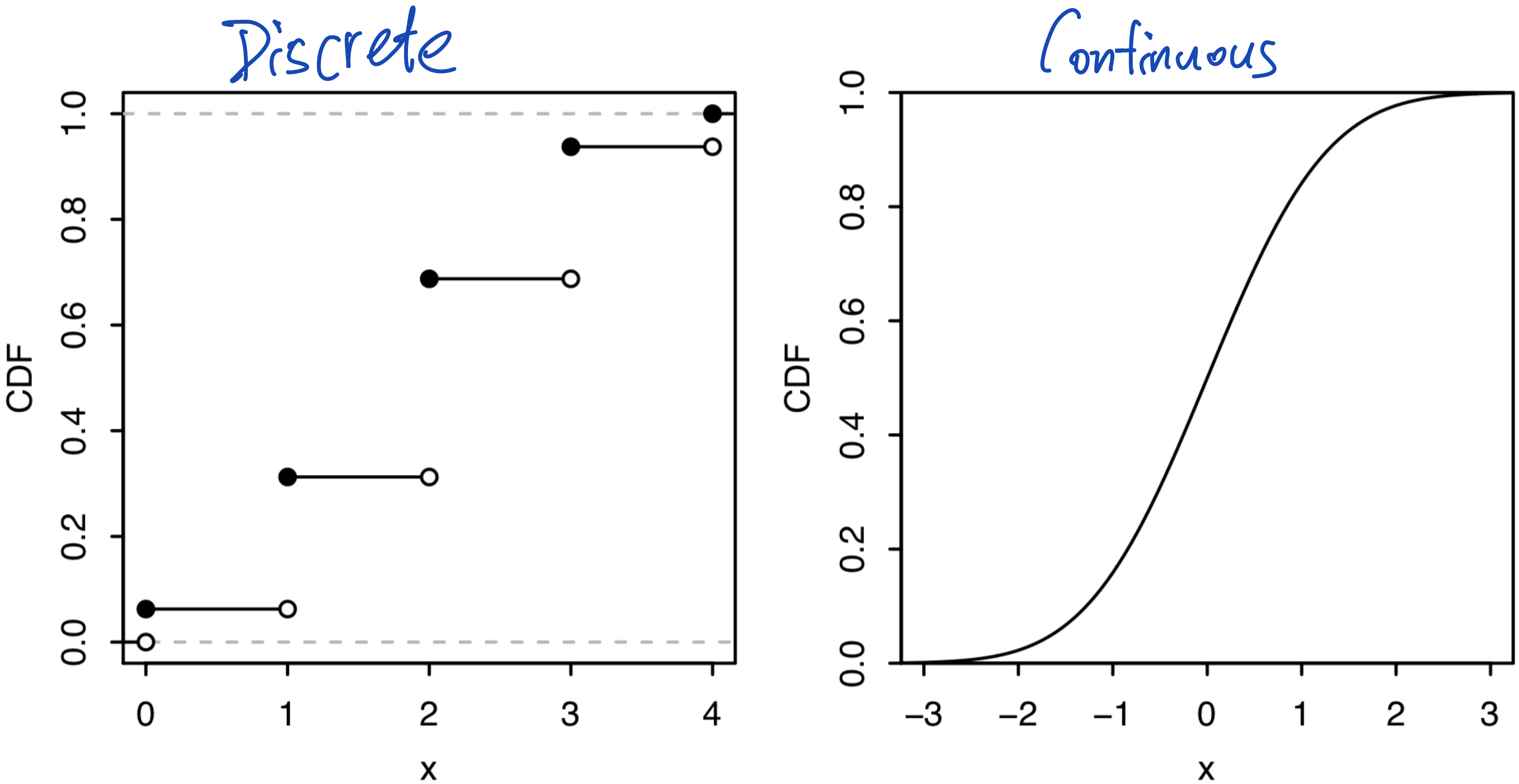
wendzh@shanghaitech.edu.cn

March 15, 2023

Outline

Outline

Discrete vs. Continuous



Continuous Random Variables

CDF可导/有限个点不可导 $\boxed{\text{可导}}$ ☆

Definition

An r.v. has a *continuous distribution* if its CDF is differentiable. We also allow there to be endpoints (or finitely many points) where the CDF is continuous but not differentiable, as long as the CDF is differentiable everywhere else. A *continuous random variable* is a random variable with a continuous distribution.

Probability Density Function

Discrete r.v. PMF

Continous r.v. PDF

Definition

For a continuous r.v. X with CDF F, the *probability density function* (PDF) of X is the derivation f of the CDF, given by $f(x) = F'(x)$. The *support* of X, and of its distribution, is the set of all x where $f(x) > 0$.

$f(x) \geq 0$? CDF increasing

Illustration of PDF

$$P(a \le X \le b) = \int_a^b f_X(x)\,dx.$$

$$= F_X(b) - F_X(a).$$

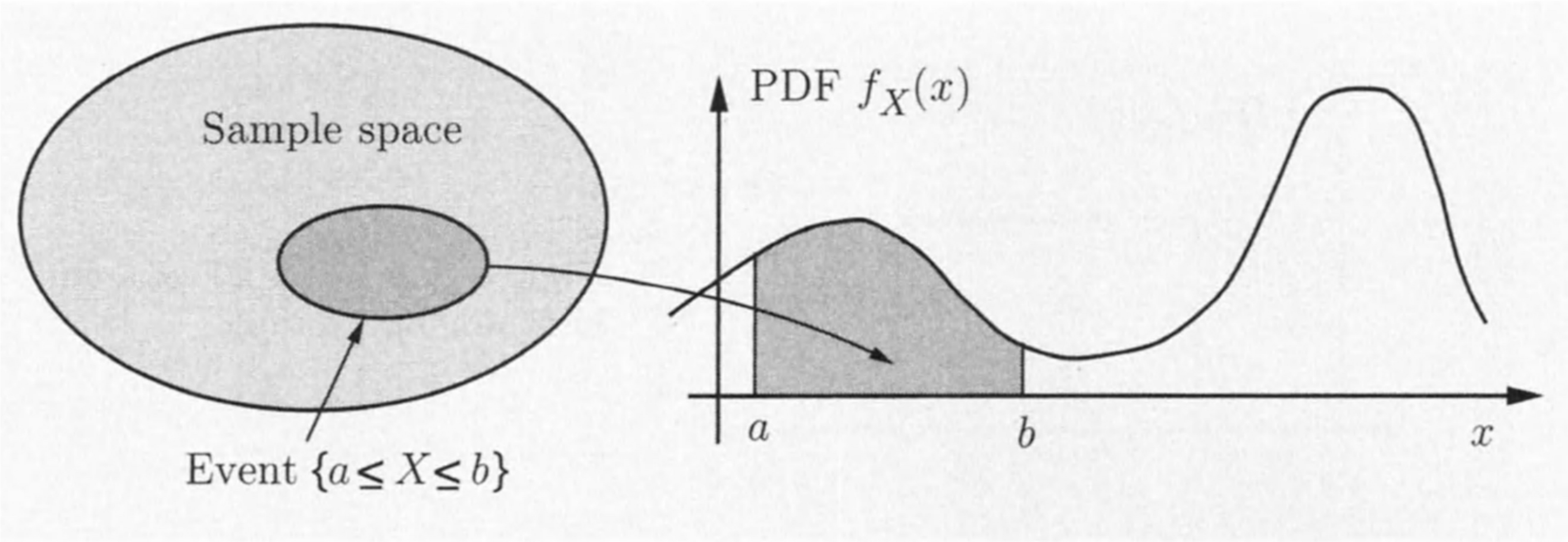
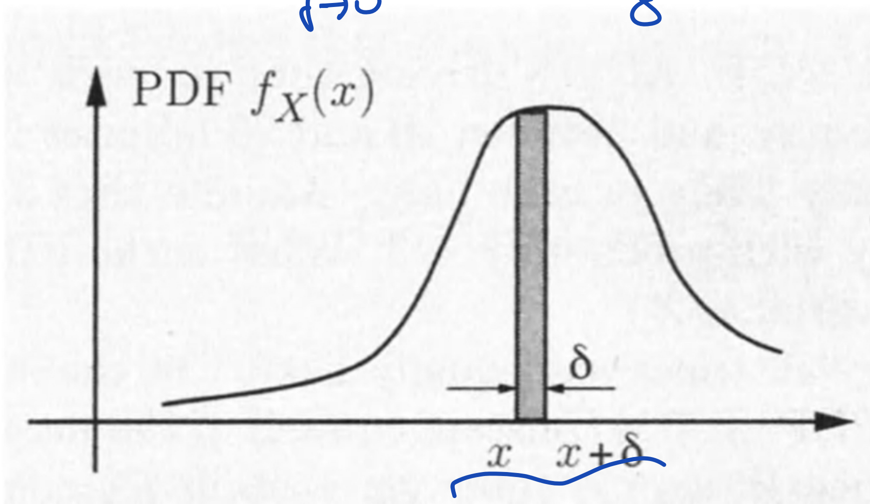


Illustration of PDF

$$\lim_{\delta\to 0} P(x \le X \le x+\delta) = \lim_{\delta\to 0}\int_x^{x+\delta} f_X(t)\,dt = f_X(t)\cdot\delta.$$

$$f_X(t) = \lim_{\delta\to 0}\frac{P(x\le X\le x+\delta)}{\delta} = \lim_{\delta\to 0}\frac{F(x+\delta)-F(x)}{\delta} = F'(x)$$



PDF vs. PMF

① PMF probability $[0,1]$

PDF $f(x) \neq$ probability. $P(A) = \int_A f(x)dx$.

$f(x) \geq 0$, $f(x) > 1$.

② PMF, X is a discrete r.v. $a \in$ Support of X.

$P(X=a) > 0$.

PDF, X is a continuous r.v. $a \in$ support of X

$P(X=a) = 0$.

$\lim_{\delta \to 0} f_X(a) \cdot \delta = 0$.

PDF to CDF

CDF $\to$ PDF $f(x) = F'(x)$

PDF $\to$ CDF $F(x) = \int_{-\infty}^{x} f(t)\,dt$

Theorem

Let X be a continuous r.v. with PDF f. Then the CDF of X is given by

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Including or Excluding Endpoints

X continuous r.v.

$$P(a<X<b) = P(a\le X<b) = P(a<X\le b) = P(a\le X\le b)$$

(a,b) $[a,b)$ $(a,b]$ $[a,b]$

$P(X=a)=0$, $P(X=b)=0$

if X, Y are continuous r.v. $P(X=Y)=?$ $0?$

could be >0

independent

$P(X=0)=0$ 确定是0，不是无穷小量

+ if X, Y are independent. $P(X=Y)=0$.

$\forall a\in\mathbb{R}$, $P(X=Y=a)=P(X=a)P(Y=a)=0$.

Valid PDFs

Theorem

The PDF f of a continuous r.v. must satisfy the following two criterias:

- *Nonnegative:* $f(x) \geq 0$; CDF montonic increasing.
- *Integrates to 1:* $\int_{-\infty}^{\infty} f(x)dx = 1$.

Example: Logistic Distribution

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

逻辑分布

The logistic distribution has CDF

$$F(x) = \frac{e^x}{1+e^x}, x \in \mathbb{R}.$$

Find the PDF.

$$f(x) = F'(x) = \frac{e^x(1+e^x) - e^{2x}}{(1+e^x)^2} = \frac{e^x}{(1+e^x)^2}$$

Example: Rayleigh Distribution

$$f(x) = F'(x) = \begin{cases} x e^{-x^2/2}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

瑞利分布

The Rayleigh distribution has CDF

$$F(x) = 1 - e^{-x^2/2}, x > 0.$$

Find the PDF.

PDF Properties

Summary of PDF properties

Let X be a continuous random variable with PDF f_X.

- $f_X(x) \geq 0$ for all x;
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$;
- If δ is very small, then $\boxed{P([x, x+\delta]) \approx f_X(x) \cdot \delta;}$
- For any subset B of the real line,

$$P(X \in B) = \int_B f_X(x)dx.$$

$a<b<c<d$

$B = [a,b] \cup (c,d)$

Expectation of a Continuous R.V.

Discrete r.v. X: $E(X) = \sum_k k \cdot P(X=k)$.

Definition

The *expected value* (also called the *expectation* or *mean*) of a continuous r.v. X with PDF f is

$$E(X) = \int_{-\infty}^{\infty} x \underbrace{\boxed{f(x)dx}}_{P(X=x)}.$$

Expectation via Survival Function

X: nonnegative integered values.

Discrete r.v. $G(n) = P(X > n)$, $E(X) = \sum_{n=0}^{\infty} G(n)$.

Theorem

Let X be a continuous and $\boxed{\textit{nonnegtive}}$ *r.v. Let F be the CDF of X, and $G(x) = 1 - F(x) = P(X > x)$. The function G is called the survival function of X. Then*

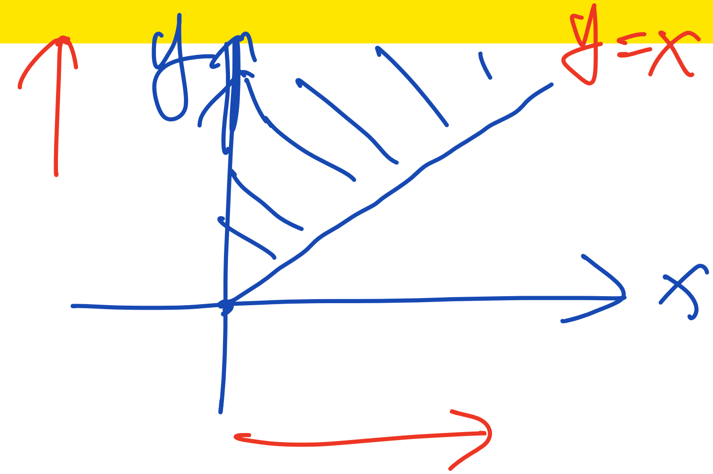
$$E(X) = \int_{\boxed{0}}^{\infty} G(x)dx$$

$\int_{x}^{+\infty} f(x)dx$

Proof

① $f_X(\cdot)$: PDF of X.

$$G(x) = P(X>x) = \int_x^\infty f_X(y)\,dy.$$



② $$\int_0^\infty G(x)\,dx = \int_0^\infty \left(\int_x^\infty f_X(y)\,dy\right)dx.$$

Fubini Theorem.

$$= \int_0^\infty \left(\int_0^y 1\,dx\right) f_X(y)\,dy.$$

$$= \int_0^\infty y\cdot f_X(y)\,dy = E(X).$$

LOTUS: Continuous

Theorem

If X is a continuous r.v. with PDF f and g is a function from $\mathbb{R}$ to $\mathbb{R}$, then

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

Symmetry Property 对称性

① $n=2$, $P(X_1<X_2)=P(X_2<X_1)=\frac{1}{2}$. $P(X_1=X_2)=0$.

② $n=3$, $P(X_1<X_2<X_3)=P(X_1<X_3<X_2)\cdots=1/6$.

i.i.d.

Continuous r.v.s that are independent and identically distributed have an important symmetry property: **all possible rankings are equally likely**.

Theorem

Let $X_1, X_2, \cdots, X_n$ be i.i.d. from a continuous distribution. Then $P(X_{a_1} < \cdots < X_{a_n}) = \frac{1}{n!}$ for any permutation $a_1, \cdots, a_n$ of $1, \cdots, n$.

$\forall i \neq j$. $P(X_i = X_j) = 0$

$\Rightarrow X_1, X_2, \cdots, X_n$ are distinct with prob. 1

$P(X_1<X_2<X_3)$

$=P(X_1<X_2, X_2<X_3)$

Proof

Proof. Let F be the CDF of X_j. By symmetry, all orderings of $X_1, \ldots, X_n$ are equally likely. For example, $P(X_3 < X_2 < X_1) = P(X_1 < X_2 < X_3)$ since both sides have exactly the same structure: they are both of the form $P(A < B < C)$ where A, B, C are i.i.d. draws from F. For any i and j with $i \neq j$, the probability of the *tie* $X_i = X_j$ is 0 since X_i and X_j are independent continuous r.v.s. So the probability of there being at least one tie among $X_1, \ldots, X_n$ is also 0, since

确定的0

$$P\left(\bigcup_{i \neq j} \{X_i = X_j\}\right) \leq \sum_{i \neq j} P(X_i = X_j) = 0.$$

Thus, $X_1, \ldots, X_n$ are distinct with probability 1, and the probability of any particular ordering is $1/n!$. ∎

Outline

Uniform Distribution 均匀分布

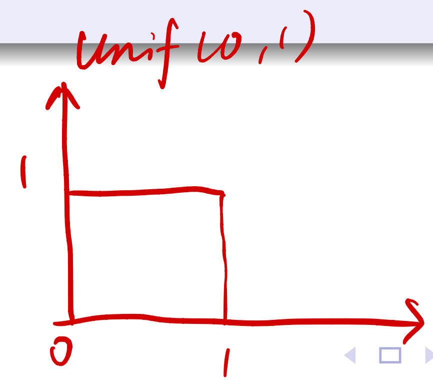
$|A| = |B|$ $\quad\quad$ $\int_A f(x)\,dx = \int_B f(x)\,dx$.

Definition

A continuous r.v. U is said to have the *Uniform Distribution* on the interval (a, b) if its PDF is

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

We denote this by $U \sim \mathrm{Unif}(a, b)$.



CDF of Uniform Distribution

CDF : $F(x) = \int_{-\infty}^{x} f(t)\,dt = P(X \le x)$.

$$= \begin{cases} 0, & \text{if } x \le a \\ \dfrac{x-a}{b-a}, & \text{if } a < x < b \\ 1, & \text{if } x \ge b. \end{cases}$$

$\int_a^x f(t)\,dt$.

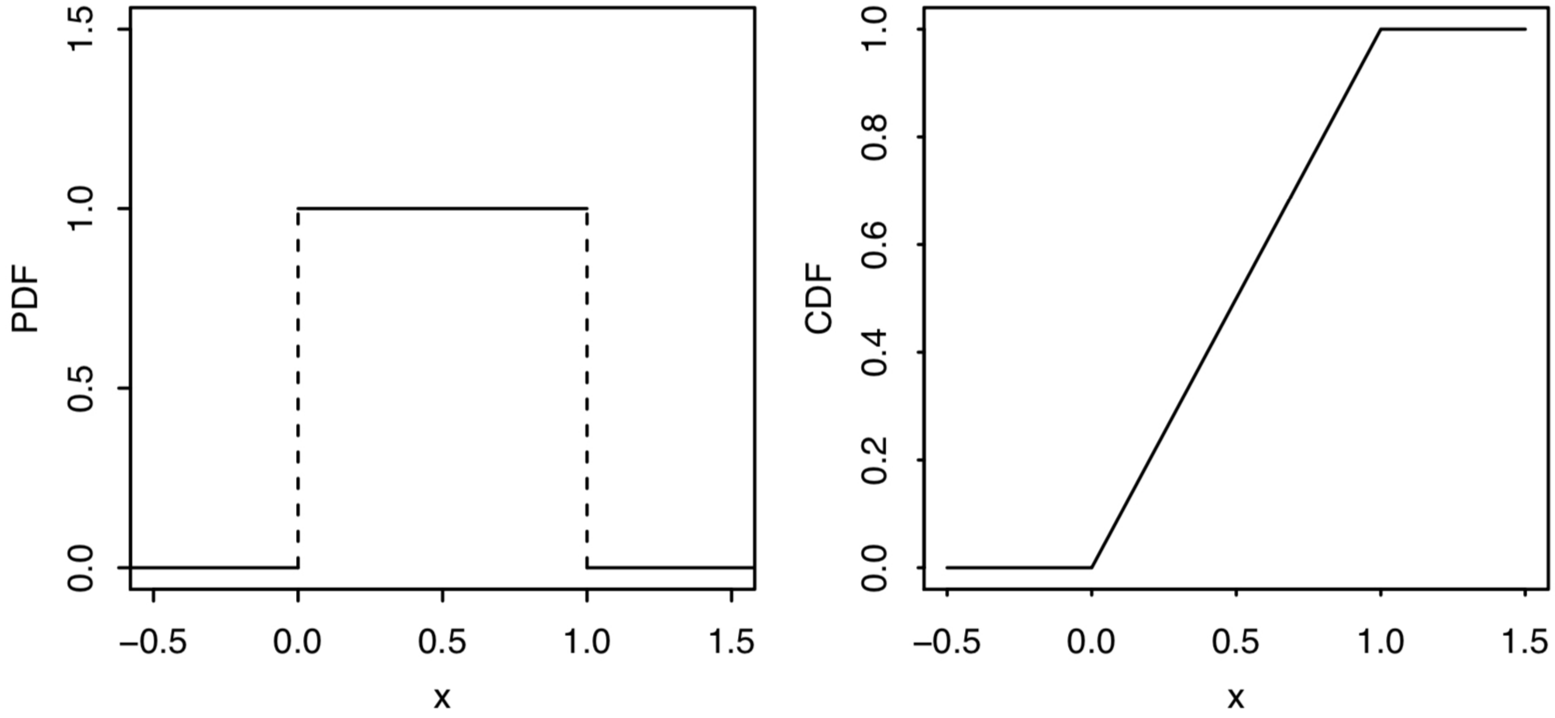
$U \sim \text{unif}(0,1)$, $a=0$, $b=1$.

CDF

$$F(t) = \begin{cases} 0, & t \le 0, \\ x, & 0 < t < x, \\ 1, & t \ge 1. \end{cases}$$

PDF & CDF

Unif(0,1)



Example

① Y. continuous r.v. $0 \le Y \le 1$.

② $y \ge 1$. $G(y) = P(Y > y) = 0$.

$0 \le y \le 1$, $G(y) = P(Y > y) = P(\min\{X_1, X_2, \cdots, X_n\} > y)$

$= P(X_1 > y)\, P(X_2 > y) \cdots P(X_n > y)$.

Suppose $X_1, X_2, \cdots, X_n$ are i.i.d $\mathrm{Unif}(0,1)$ random variables and let $Y = min(X_1, X_2, \cdots, X_n)$ be their minimum. Find $E(Y)$.

$E(Y) = \int_{0}^{+\infty} G(y)\,dy$

$= [P(X_1 > y)]^n = [1 - F_{X_1}(y)]^n$

$= (1-y)^n$

③ $E(Y) = \int_0^{\infty} G(y)\,dy = \int_0^1 G(y)\,dy = \int_0^1 (1-y)^n\,dy = \frac{1}{n+1}$.

Outline

Universality of the Uniform

- Given a $\mathrm{Unif}(0,1)$ r.v., we can construct an r.v. with any continuous distribution we want.
- Conversely, given an r.v. with an arbitrary continuous distribution, we can create a $\mathrm{Unif}(0,1)$ r.v.
- Other names:
    - probability integral transform
    - inverse transform sampling
    - the quantile transformation
    - the fundamental theorem of simulation

普适性

Lemma: $U \sim unif(0,1)$ $P(U \le y) = y$. if $y \in [0,1]$

Universality of the Uniform

① $U \sim unif(0,1)$, $X = F^{-1}(U)$; CDF of X: $F_X(x)$.

$F_X(x) = P(X \le x) = P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x)$, $\forall x$.

$0 \le F(x) \le 1$

support 连续，严格单调 $f(x) > 0$

Theorem

Let F be a CDF which is a continuous function and strictly increasing on the support of the distribution. This ensures that the inverse function F^{-1} exists, as a function from $(0,1)$ to $\mathbb{R}$. We then have the following results.

1. *Let $U \sim \mathrm{Unif}(0,1)$ and $X = F^{-1}(U)$. Then X is an r.v. with CDF F.*
2. *Let X be an r.v. with CDF F. then $F(X) \sim \mathrm{Unif}(0,1)$.*

support $x: f(x) > 0$

② $Y = F(X) \in [0,1]$, $P(Y \le y) = 0$, $y \le 0$.

$P(Y \le y) = 1$, if $y \ge 1$.

CDF↑

$\Rightarrow F^{-1}(u) \le x \Leftrightarrow u \le x$

Proof: Universality of the Uniform

$y \in (0,1)$. $\quad P(Y \le y) = P(F(X) \le y)$

$= \underbrace{P(X \le F^{-1}(y))}$ $\quad F_X(x) = P(X \le x)$

$= F(F^{-1}(y))$

$= y$.

$$P(Y \le y) = F_Y(y) = \begin{cases} 0, & y \le 0 \\ y, & 0 < y < 1, \\ 1, & y \ge 1. \end{cases} \sim \mathrm{Unif}(0,1)$$

Example: Universality with Logistic

① CDF of logistic distribution: $F(x) = \frac{e^x}{1+e^x}$, $x \in R$.

② We have $U \sim Unif(0,1)$, if we want to get a r.v. logistic.

$y = \frac{e^x}{e^x+1}$

$F^{-1}(x) = \log \frac{x}{1-x}$, $x \in (0,1)$

$e^x(y-1) = -y$ $e^x = \frac{y}{1-y}$ $x = \ln\frac{y}{1-y}$

$\Rightarrow F^{-1}(x) = \log\frac{x}{1-x}$

③ $F^{-1}(U) = \log\left(\frac{U}{1-U}\right) \sim$ logistic distribution

verify. CDF: $P\left[\log\left(\frac{U}{1-U}\right) \leq x\right] = P\left[\frac{U}{1-U} \leq e^x\right]$

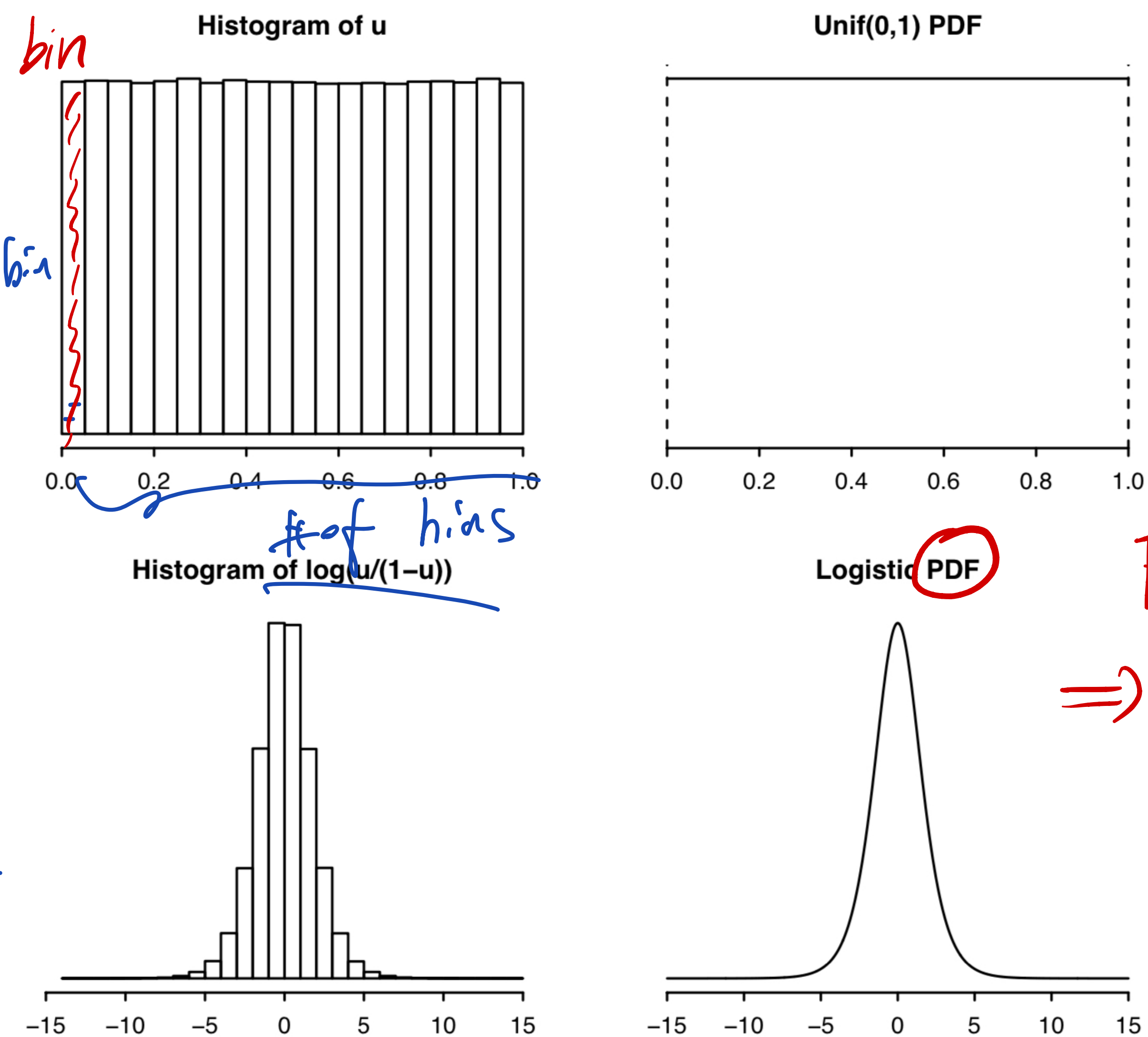
$$= P\Big[U \leq \underbrace{\frac{e^x}{1+e^x}}_{\in(0,1)}\Big] = \frac{e^x}{1+e^x} = F(x).$$

Histogram 直方图

- Introduced by Karl Pearson.
- A graphical representation of the distribution of numerical data.
- An estimate of the probability distribution (density estimation) of a continuous variable.
- To construct a histogram, the first step is to "bin" the range of values: divide the entire range of values into a series of intervals and then count how many values fall into each interval.
- The bins are usually specified as consecutive, non-overlapping intervals of a variable.

Histogram & PDF

U



$F^{-1}(U)$

$F^{-1}(U)$

$\Rightarrow$ F对应的 PDF

Example: Universality with Rayleigh

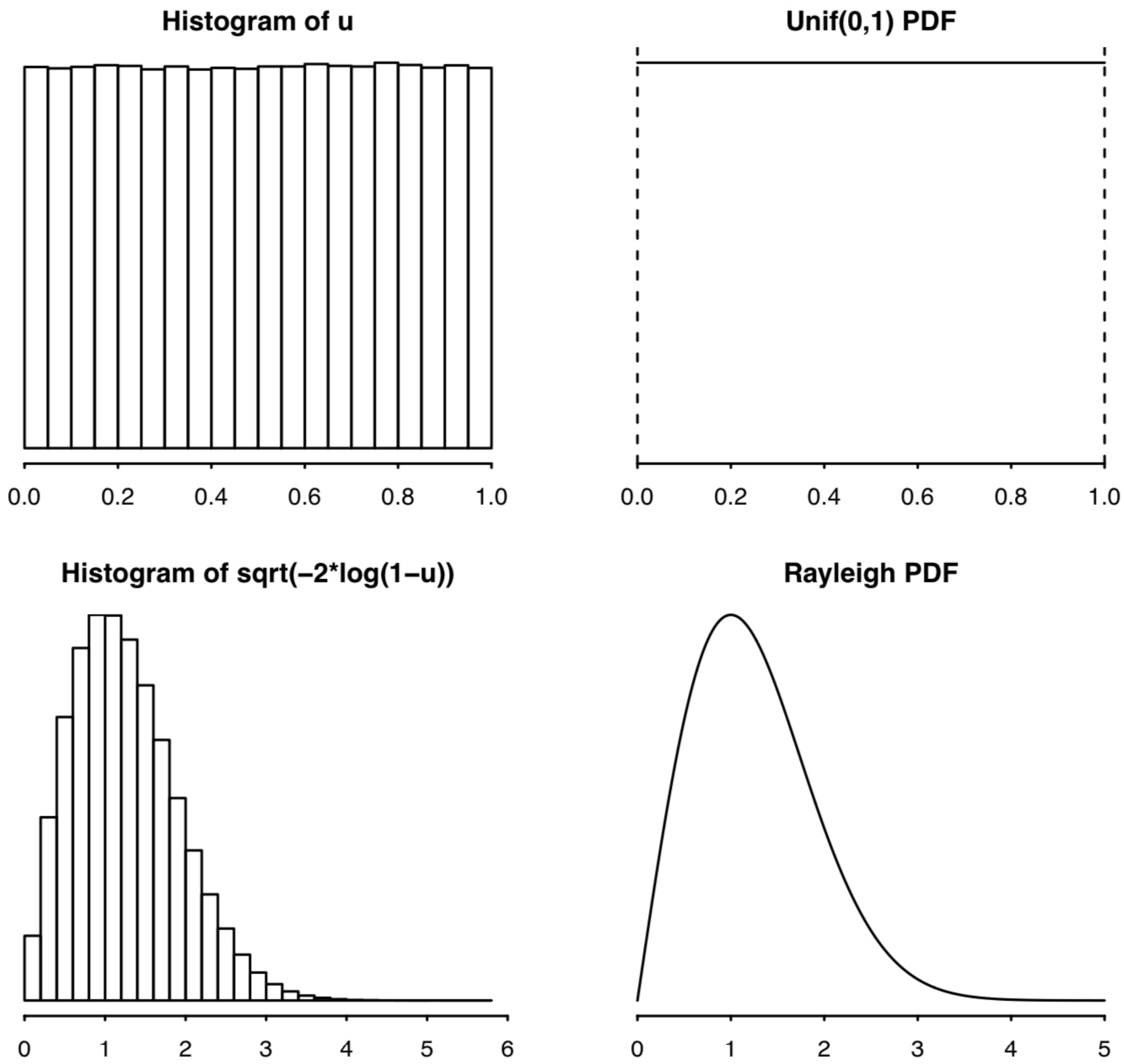
① $F(x) = 1 - e^{-\frac{1}{2}x^2}$, $x > 0$. CDF of Rayleigh.

② $F^{-1}(x) = \sqrt{-2\log(1-x)}$, $0 < x < 1$.

③ $U \sim \text{unif}(0,1)$.

$F^{-1}(U) = \sqrt{-2\log(1-U)}$, $\sim$ Rayleigh distribution.

Histogram & PDF



Outline

Exponential Distribution

指数分布

Definition

A continuous r.v. X is said to have the *Exponential Distribution* with parameter λ if its PDF is

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

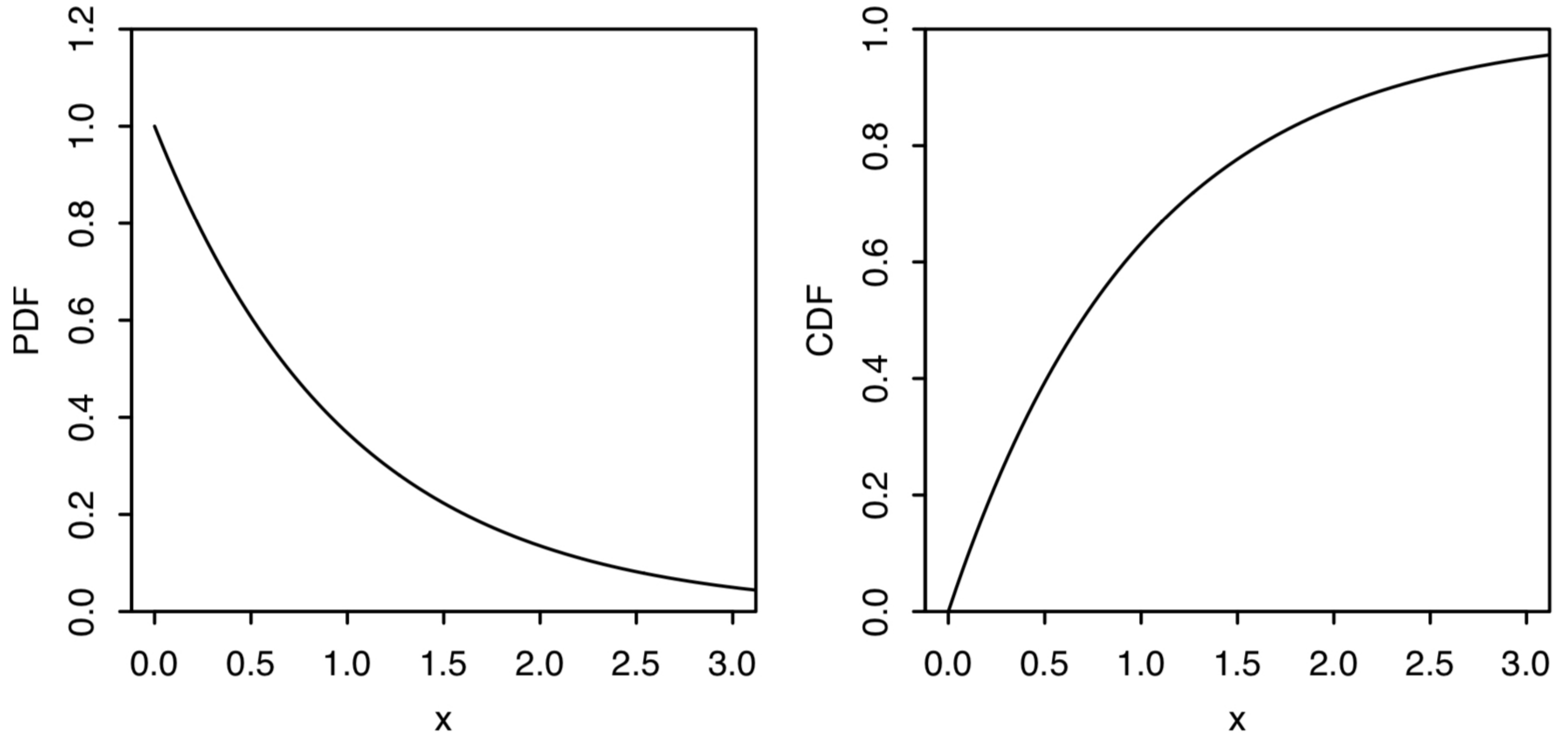
We denote this by $X \sim \mathrm{Expo}(\lambda)$. The corresponding CDF is

$$F(\cancel{X}^{x}) = 1 - e^{-\lambda x}, x > 0.$$

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(\tau)d\tau = \int_0^x \lambda e^{-\lambda\tau}d\tau = -e^{-\lambda\tau}\Big|_{\tau=0}^{x} = -(e^{-\lambda x}-1) = 1-e^{-\lambda x}$$

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

Expo(1) PDF & CDF



Memoryless Property

X nonegative integered value.

Discrent r.v. X. $P(X \geq n+k \mid X \geq k) = P(X \geq n)$. memoryless property

Continuous r.v. $X \sim \mathrm{Expo}(\lambda) \Rightarrow P(X \geq t) = P(X > t) = e^{-\lambda t}$

无记忆性质

Definition

A distribution is said to have the *memoryless property* if a random variable X from that distribution satisfies

$$P(X \geq s+t \mid X \geq s) = P(X \geq t) = P(X > t) = 1 - P(X \leq t) = e^{-\lambda t}.$$

for all $s, t > 0$.

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

Memoryless Property

Minimum of independent Expos

$X_j \sim \mathrm{Expo}(\lambda_j),\ P(X_j > t) = e^{-\lambda_j t}$

① $\forall t > 0,\ P(L > t) = P(\min(X_1, X_2, \cdots, X_n) > t)$

$= P(X_1 > t, X_2 > t, \cdots, X_n > t).$

$= P(X_1 > t) P(X_2 > t) \cdots P(X_n > t).$

$= e^{-\lambda_1 t} \ e^{-\lambda_2 t} \cdots e^{-\lambda_n t}$

Let $X_1, \cdots, X_n$ be independent, with $X_j \sim \mathrm{Expo}(\lambda_j)$. Let $L = \min(X_1, \cdots, X_n)$. Show that $L \sim \mathrm{Expo}(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$, and interpret this intuitively.

$= e^{-(\lambda_1 + \lambda_2 + \cdots + \lambda_n) t} \Rightarrow L \sim \mathrm{Expo}(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$

Failure (Hazard) Rate Function

① $r(t) = \lim_{dt \to 0} \frac{P(X \in [t, t+dt] \mid X > t)}{dt}$

② $\lim_{dt \to 0} \frac{P(X \in (t, t+dt) \mid X > t)}{dt} = \lim_{dt \to 0} \frac{P(X \in (t, t+dt), X > t)}{P(X > t)\, dt} = \lim_{dt \to 0} \frac{P(X \in (t, t+dt))}{dt(1-F(t))}$

失效率函数

$r(t)\uparrow$ 衰减越快

Definition

Let X be a continuous random variable with PDF $f(t)$ and CDF $F(t) = P(X \leq t)$. Then the failure (hazard) rate function $r(t)$ is

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

$= \frac{f(t) \cdot \cancel{dt}}{(1 - F(t))\cancel{dt}}$

- $r(t)$: an instantaneous failure rate of t-year-old-item,
- $r(t) = \lambda$ for exponential distribution with parameter λ.

$X \sim Expo(\lambda)$. $f(t) = \lambda e^{-\lambda t}$ $F(t) = 1 - e^{-\lambda t}$. $r(t) = \frac{f(t)}{1-F(t)} = \lambda$.

衰减速率为常数 ⇒ 指数衰减

Why Exponential Distribution

- Some physical phenomena, such as radioactive decay, truly do exhibit the memoryless property.
- The Exponential distribution is well-connected to other named distributions (Poisson distribution)
- The Exponential serves as a building block for more flexible distributions, such as the Weibull distribution, that allow for a wear-and-tear effect (where older units are due to break down) or a survival-of-the-fittest effect (where the longer you've lived, the stronger you get).

Memoryless Property

① Memoryless: $P(X \geq s+t \mid X \geq s) = P(X \geq t) \Rightarrow \dfrac{P(X \geq s+t)}{P(X \geq s)} = P(X \geq t)$

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

② $G(x) = P(X > x) = P(X \geq x)$

$P(X > s+t) = P(X > s)P(X > t)$

$G(s+t) = G(s)G(t)$

Theorem

If X is a positive continuous random variable with the memoryless property, then X has an Exponential distribution.

无记忆性质 $\Leftrightarrow$ 指数分布

③ $G(s+t) = G(s)G(t)$

Proof

$G(t) = G(\frac{t}{2} + \frac{t}{2}) = G^2(\frac{t}{2})$

$s = t \Rightarrow G(2t) = G^2(t),\ G(3t) = G^3(t),\ \ldots,\ G(kt) = G^k(t);$

$G(\frac{t}{2}) = G^{1/2}(t),\ G(\frac{t}{3}) = G^{1/3}(t),\ \ldots,\ G(\frac{t}{k}) = G^{\frac{1}{k}}(t),$

$G(\frac{m}{n}t) = G^{\frac{m}{n}}(t)\quad \forall m, n \in \mathbb{Z},\ n \neq 0.$

1) $\{\frac{m}{n}\}$ includes all rational numbers.

2) For irrational numbers, it can be regarded as the limit of a sequence of rational numbers.

3) Since $G(x)$ is continuous, $G(x) = G(1)^x = e^{x \ln G(1)}$

$\forall x \in \mathbb{R}.$

Geometric Distribution is also Memoryless

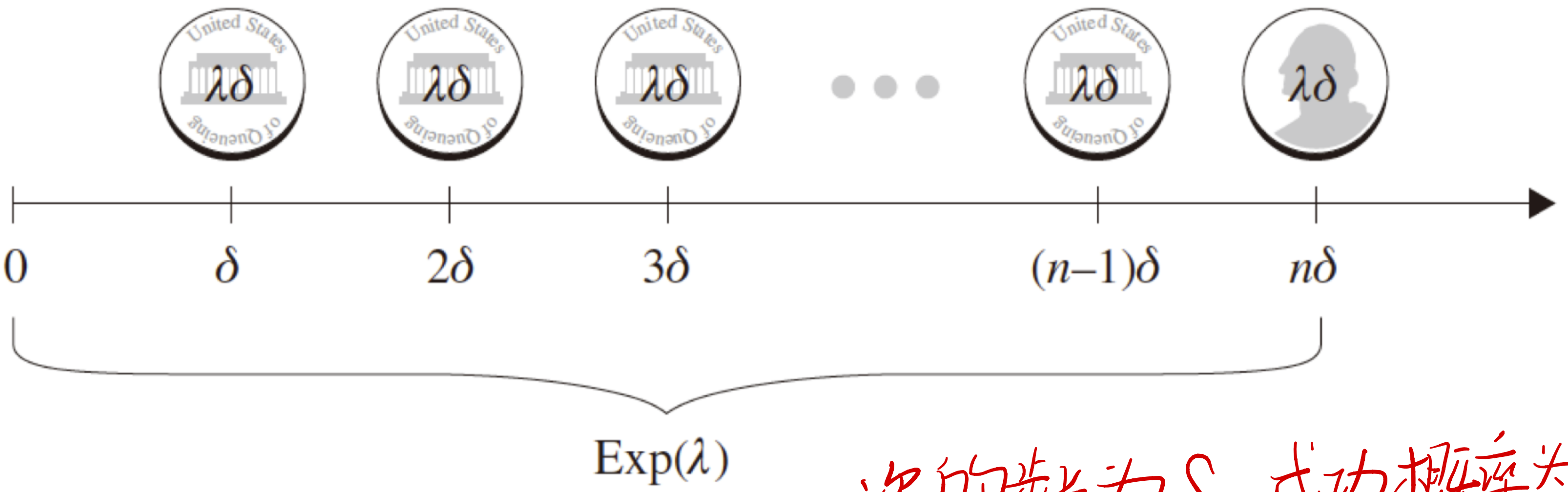
Geom(p), FS(p) ⇒ memoryless
⇒ Expo(·)

- Exponential distribution as the "continuous counterpart" of the Geometric distribution (or First Success Distribution).
- Recall that the First Success Distribution can be viewed as the number of flips needed to get a "success".
- The distribution of the remaining number of flips is independent of how many times we have flipped so far.
- The same holds for the Exponential distribution, which is the time until "success".

Connecting Discrete and Continuous: δ-Steps Method

- Very useful tools to develop intuition.
- Connecting Geometric and Exponential Distributions.
- Connecting DTMC (discrete-time Markov chain) CTMC (continuous-time Markov chain).
- Also applied to Poisson process.

Exponential & Geometric via δ-Steps



一次的步长为δ，成功概率为$\lambda\delta$

- An Exponential random variable with λ represents the time to a successful event, given that an event occurs every δ-step and is successful with probability $\lambda\delta$, where $\delta \to 0$.

Exponential & Geometric via δ-Steps

① Divide one time unit into n intervals, each of which has a length of $\delta = \frac{1}{n}$.

Trails occur at the beginning of every interval, each trail with success prob. $\lambda\delta$.

② Define a r.v. Y: # of trails until the first success.

Y is a discrete r.v.: $Y \sim FS(\lambda\delta \mid \text{flip every } \delta\text{-step})$

$$E(Y) = \frac{1}{\lambda\delta}$$

③ Define a r.v. $\hat{Y}$: the time until the first success event.

$\hat{Y}$ is a continuous r.v. $\hat{Y} = Y \cdot \delta$ 次数为Y, 一次时长δ

$$E(\hat{Y}) = E(Y)\cdot\delta = \frac{1}{\lambda}.$$

Exponential & Geometric via δ-Steps

④ Consider the survival function of $\hat{Y}$, $\forall t > 0$. integer

$$P(\hat{Y} > t) = P(Y \cdot \delta > t) = P(Y > \frac{t}{\delta}) = P(\underline{Y-1} \geq \underline{\underline{\frac{t}{\delta}}}) = (1-\lambda\delta)^{\frac{t}{\delta}}$$

$Y \sim FS(\lambda\delta)$. $Y-1 \sim Geom(\lambda\delta)$.

Lemma: $X \sim Geom(p)$, $P(X \geq n) = (1-p)^n$.

⑤ as $\delta \to 0$ $(n \to \infty)$, $\hat{Y} \sim Expo(\lambda)$

$$\lim_{\delta \to 0} (1-\lambda\delta)^{\frac{t}{\delta}} = \lim_{\delta \to 0} e^{\frac{t \ln(1-\lambda\delta)}{\delta}} = e^{-\lambda t}$$

t/δ, not integer.

$$P(Y > \frac{t}{\delta}) = P(Y > \lfloor \frac{t}{\delta} \rfloor)$$

$$= P(Y-1 \geq \lfloor \frac{t}{\delta} \rfloor) = \underline{(1-\lambda\delta)^{\lfloor \frac{t}{\delta} \rfloor}}, \to e^{-\lambda t}$$

$\frac{t}{\delta} - 1 < \lfloor \frac{t}{\delta} \rfloor < \frac{t}{\delta}$, $\delta \to 0$

Story: Connect Exponential with Poisson

2辆车到达的时间间隔

Poisson process

任意时间t内，到达数 $\sim Pois(\lambda t)$

Definition

A process of arrivals in continuous time is called a *Poisson process* with rate λ if the following two conditions hold.

$P(N_t = n) = \frac{e^{-\lambda t} \cdot (\lambda t)^n}{n!}, \quad n \geq 0.$

1. The number of arrivals that occur in an interval of length t is a $\mathrm{Pois}(\lambda t)$ random variable.

   N_t: # of arrivals within time t, $N_t \sim \mathrm{Pois}(\lambda t)$

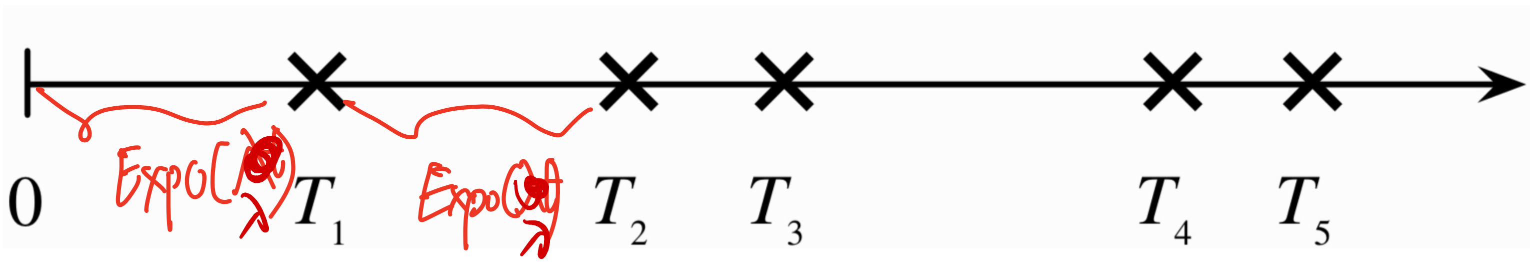
2. The numbers of arrivals that occur in disjoint intervals are independent of each other. For example, the numbers of arrivals in the intervals $(0, 10)$, $[10, 12)$, and $[15, \infty)$ are independent.

Story: Connect Exponential with Poisson

① $N(t)$: # of arrivals within $[0,t]$; $N(0)=0$;

T_n = the n-th arrival time

$N(t) = \max_n \{T_n \le t\}$; $P(N(t)=n) = \frac{e^{-\lambda t} \cdot (\lambda t)^n}{n!}$, $n \ge 0$.



② Duality: $\{N(t) \ge n\} \Leftrightarrow \{T_n \le t\}$.

$\{T_1 > t\} \Leftrightarrow \{N(t) = 0\}$. $P(T_1 > t) = P(N(t)=0) = e^{-\lambda t}$, $\forall t$.

$T_1 \sim \text{Expo}(\lambda)$

$T_n = \sum_{i=1}^{n} X_n$.

③ $X_i = \underline{T_i - T_{i-1}}$, $i \ge 1$. $X_i \sim$ i.i.d Expo(λ),

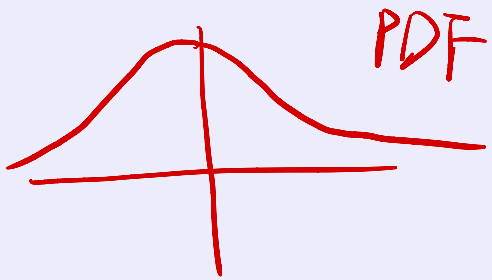
Outline

Standard Normal Distribution

标准正态分布

Definition

A continuous r.v. Z is said to have the *standard Normal distribution* if its PDF φ is given by



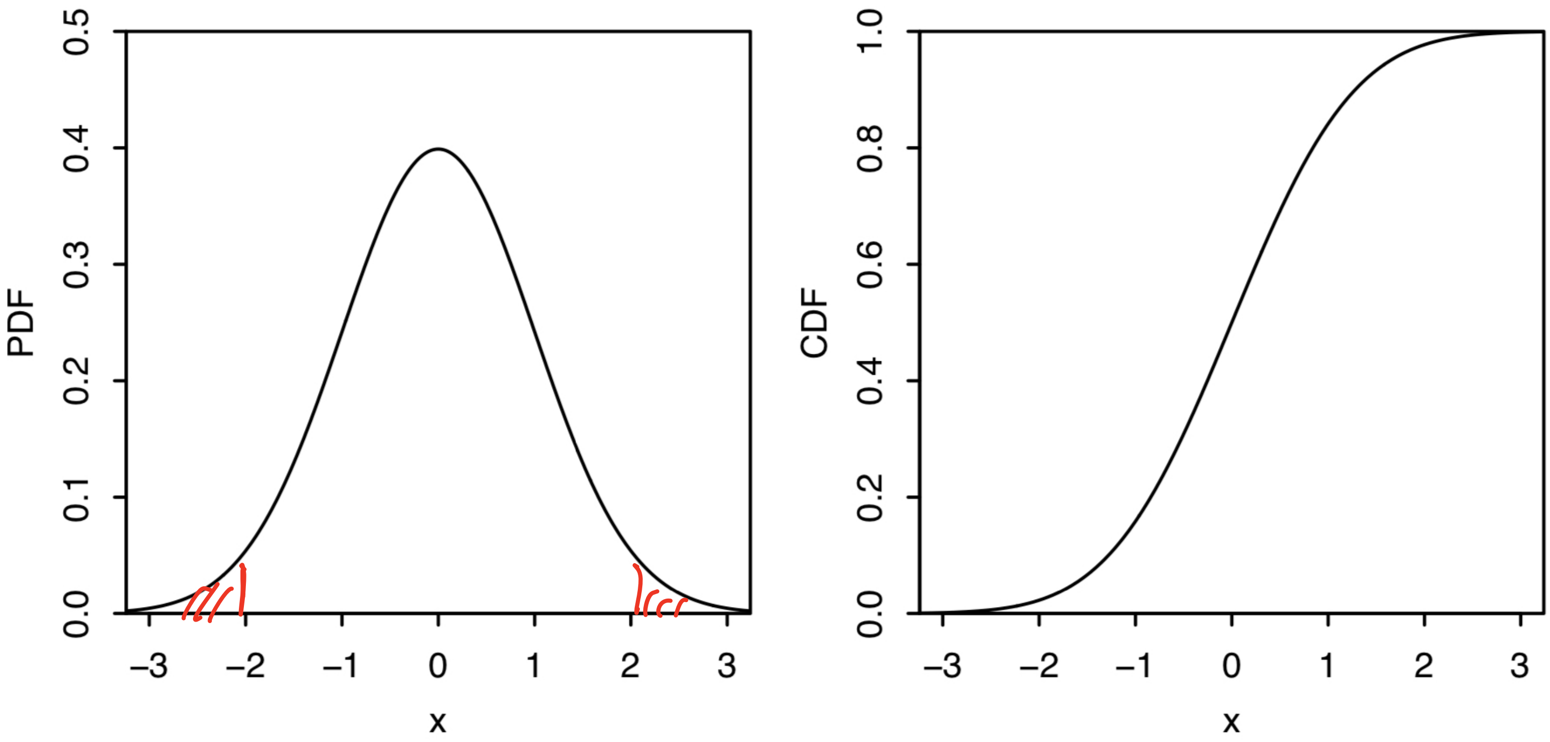
$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, -\infty < z < \infty.$$

We write this as $Z \sim \mathcal{N}(0, 1)$ since, as we will show, Z has mean 0 and variance 1.

The standard Normal CDF Φ is the accumulated area under the PDF:

$$\Phi(z) = \int_{-\infty}^{z} \varphi(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

PDF & CDF



$\Phi(-2) = 1 - \Phi(2)$

Property of Standard Normal PDF & CDF

- φ for the standard Normal PDF, Φ for the CDF and Z for the r.v.
- Symmetry of PDF: $\varphi(z) = \varphi(-z)$.
- Symmetry of tail areas: $\Phi(z) = 1 - \Phi(-z)$. 关于 $x=0$ 对称、
- Symmetry of Z and $-Z$: If $Z \sim \mathcal{N}(0,1)$, then $-Z \sim \mathcal{N}(0,1)$.
- Mean is 0 and variance is 1.

Varify the Validity of PDF

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 1 ?$$

① $$\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz\right)^2 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)} dxdy = \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{-\frac{1}{2}r^2} r dr d\theta = 1$$

$$\begin{cases} x = r\cos\theta, \\ y = r\sin\theta \end{cases} \qquad dxdy = rdrd\theta$$

Mean & Variance

$E(Z) = 0, \quad Var(Z) = 1. = E(X^2) - [E(X)]^2$

Normal Distribution

$E(X) = \mu + E[\sigma Z] = \mu + \sigma \cdot E(Z) = \mu$

$Var(X) = Var(\mu + \sigma Z) = Var(\sigma Z)$

$= \sigma^2 Var(Z) = \sigma^2$

Definition

If $Z \sim \mathcal{N}(0, 1)$, then

standard

$$\boxed{X = \mu + \sigma Z}$$

linear transformation.

is said to have the *Normal distribution* with mean μ and variance σ^2. We denote this by $X \sim \mathcal{N}(\mu, \sigma^2)$.

Normal CDF and PDF

$F(x) = P(X \le x)$
$= P(\mu + \sigma Z \le x)$
$= P(Z \le \frac{x-\mu}{\sigma}) = \Phi(\frac{x-\mu}{\sigma})$.
$X = \mu + \sigma Z$
$\Downarrow$

Theorem

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then the CDF of X is $Z = \frac{X-\mu}{\sigma}$

$$\boxed{F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)},$$

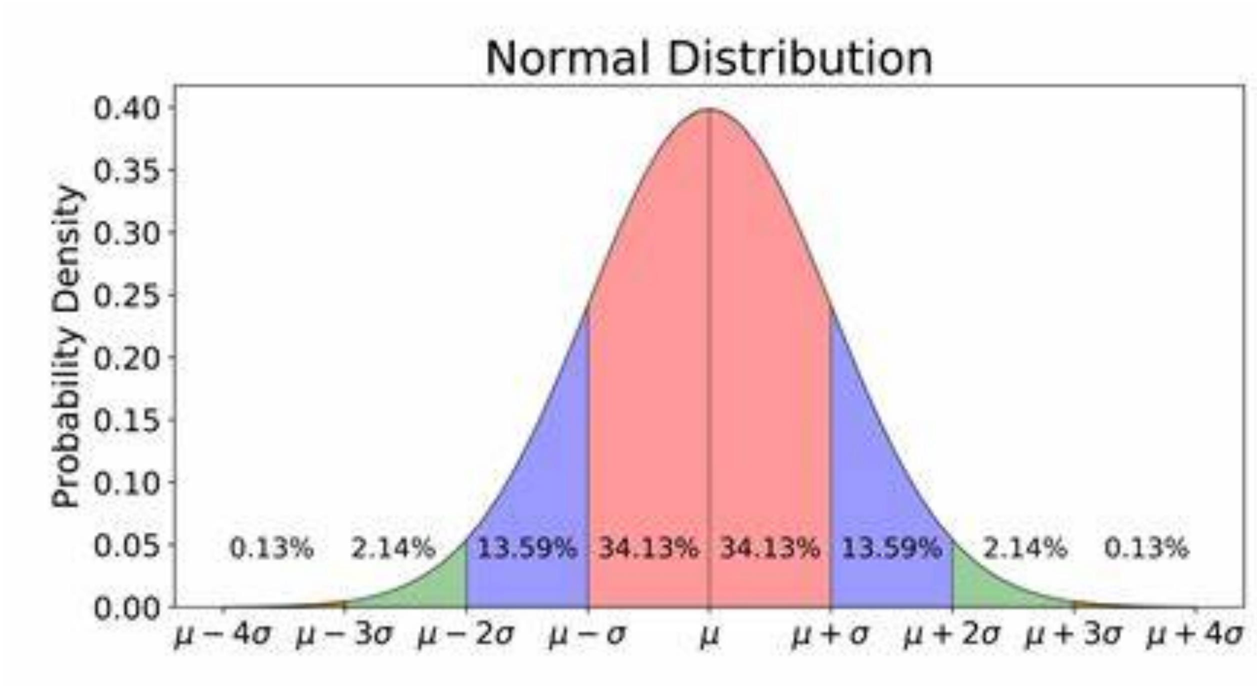
$\varphi(x) = \Phi'(x)$

and the PDF of X is

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

$f(x) = F'(x)$

$68 - 95 - 99.7\%$ Rule



$$P(\mu - 1\cdot\sigma \le X \le \mu + 1\cdot\sigma) \approx 68.27\%$$
$$P(\mu - 2\cdot\sigma \le X \le \mu + 2\cdot\sigma) \approx 95.45\%$$
$$P(\mu - 3\cdot\sigma \le X \le \mu + 3\cdot\sigma) \approx 99.73\%$$

Outline

Sample Mean

$E(\bar{X}_n) = E[\frac{1}{n}\sum_{j=1}^{n} X_n] = \frac{1}{n}\sum_{j=1}^{n} E(X_n) = \mu.$

$n\to\infty$个 iid $\bar{X}_n$ 最终收敛于 $N(\mu, \frac{\sigma^2}{n})$

Definition

Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables with finite mean μ and finite variance σ^2. The *sample mean* $\bar{X}_n$ is defined as follows:

$$\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

The sample mean $\bar{X}_n$ is itself an r.v. with mean μ and variance σ^2/n.

$Var(\bar{X}_n) = \frac{1}{n^2}\cdot Var(\sum_{j=1}^{n} X_j) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.$

$n\to\infty$ $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$

Central Limit Theorem

中心极限定理

$\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$.

$$\frac{\bar{X}_n - \mu}{\frac{1}{\sqrt{n}}\sigma}$$

Theorem

As $n \to \infty$,

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

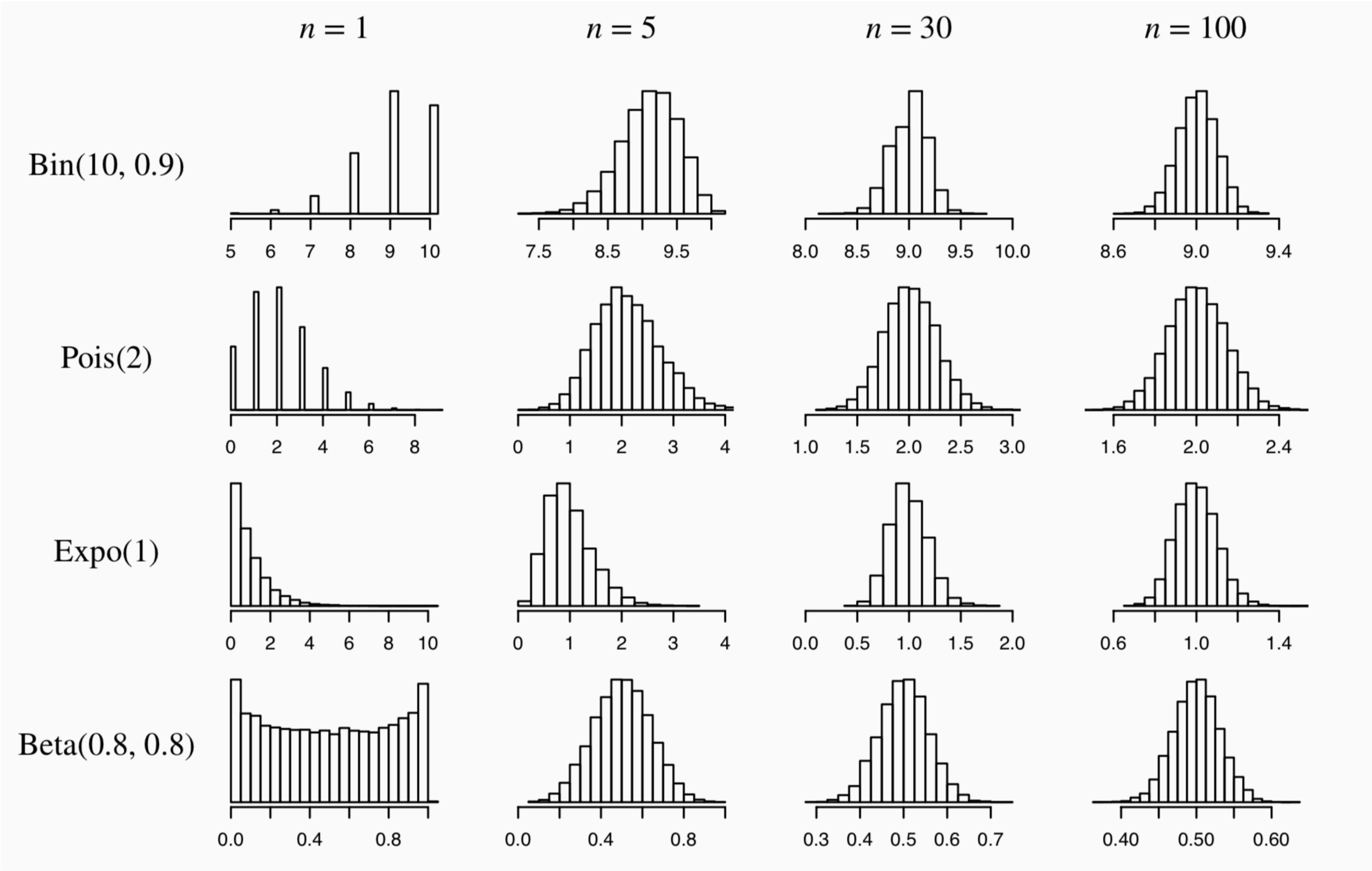
In words, the CDF of the left-hand side approaches the CDF of the standard Normal distribution.

$\bar{X}_n = \mu + \frac{\sigma}{\sqrt{n}} Z$ $\qquad$ $Z = \left(\frac{\bar{X}_n - \mu}{\sigma}\right)\sqrt{n}$

CLT Approximation

- For large n, the distribution of $\bar{X}_n$ is approximately $\mathcal{N}(\mu, \sigma^2/n)$.
- For large n, the distribution of $n\bar{X}_n = X_1 + X_2 + \cdots + X_n$ is approximately $\mathcal{N}(n\mu, n\sigma^2)$.

CLT Approximation: Example



Poisson Convergence to Normal

$X \sim Pois(\lambda_1)$, $Y \sim Pois(\lambda_2)$. X, Y independent

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

Let $Y \sim \mathrm{Pois}(n)$. We can consider Y to be a sum of n i.i.d. $\mathrm{Pois}(1)$ r.v.s. Therefore, for large n,

$$Y \sim \mathcal{N}(n, n)$$

$X_i \sim Pois(\lambda_i)$

$\sum_{i=1}^{n} X_i \sim Pois(\lambda_1 + \cdots + \lambda_n)$

Binomial Convergence to Normal

Let $Y \sim \mathrm{Bin}(n, p)$. We can consider Y to be a sum of n i.i.d. $\mathrm{Bern}(p)$ r.v.s. Therefore, for large n,

$$Y \sim \mathcal{N}\left(np, np(1-p)\right).$$

Continuity Correction: De Moivre-Laplace Approximation

$Y \sim Bin(n, p)$

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

- Poisson approximation: when n is large and p is small, $np = \lambda$ constant
- Normal approximation: when n is large and p is around $1/2$.

De Moivre-Laplace Approximation

$$P(k \leq Y \leq l) = P(k - \frac{1}{2} < Y < l + \frac{1}{2})$$
$$\approx \Phi\left(\frac{l + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - 1/2 - np}{\sqrt{np(1-p)}}\right).$$

- Very good approximation when $n \leq 50$ and p is around $1/2$.

Example

Let $Y \sim \mathrm{Bin}(n, p)$ with $n = 36$ and $p = 0.5$.

$Y = \sum_{k=0}^{21} P(Y=k)$.

- An exact calculation: $P(Y \leq 21) = 0.8785$

$Y = \sum_{i=1}^{n} Y_i$. $Y_i \sim \mathrm{Bern}(\frac{1}{2})$.

- CLT approximation:

Y Approximately follows $N(np, np(1-p))$

$$P(Y \leq 21) \approx \Phi\left(\frac{21 - np}{\sqrt{np(1-p)}}\right) = \Phi(1) = 0.8413$$

> 0.03

- DML approximation:

$$P(Y \leq 21) \approx \Phi\left(\frac{21.5 - np}{\sqrt{np(1-p)}}\right) = \Phi(1.17) = 0.879$$

0.0005

History

- 1733: normal distribution was introduced by French mathematician Abraham DeMoivre.
- Abraham DeMoivre (1667–1754): worked at betting shop, computing the probability of gambling bets in all types of games of chance. Also a close friend of Isaac Newton.
- 1809: rediscovered by German mathematician Karl Friedrich Gauss, and then people call it the Gaussian distribution.

History

- During the mid-to-late 19th century, most statisticians started to believe that the majority of data sets would have histograms conforming to the Gaussian bell-shaped form.
- Indeed, it came to be accepted that it was “normal” for any well-behaved data set to follow this curve.
- Following the lead of the British statistician Karl Pearson, we also call “normal distribution”.

Family of Normal Distribution

- Chi-Square Distribution: Found by Karl Pearson.
- Student-t Distribution: Found by Student (William Gosset).
- F-distribution: Found by Ronald Fisher.

Family of Normal Distribution

Given i.i.d. r.v.s
$X_i \sim \mathcal{N}(0,1),\ Y_j \sim \mathcal{N}(0,1),\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, m$. Then we have

- Chi-Square Distribution: χ^2分布 (卡方)

$$\chi_n^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

- Student-t Distribution:

$$t = \frac{Y_1}{\sqrt{\frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n}}}$$

- F Distribution:

$$F = \frac{\frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n}}{\frac{Y_1^2 + Y_2^2 + \cdots + Y_m^2}{m}}$$

Outline

Moment Generating Function 矩生成函数

$$M(t) = \begin{cases} \sum_k e^{tk} P(X=k), & \text{discrete variables. (PMF),} \\ \int_{-\infty}^{+\infty} e^{tx} f_X(x)\,dx, & \text{continuous variables, (PDF).} \end{cases}$$

Definition

The *moment generating function* (MGF) of an r.v. X is $M(t) = E(e^{tX})$, as a function of t, if this is finite on some open interval $(-a, a)$ containing 0. Otherwise we say the MGF of X does not exist

$M_X(0) = E(e^{t \cdot 0}) = 1$

存在一个包含0的对称区间

s.t. $M(t)$有限

Bernoulli MGF

$X \sim \text{Bern}(p)$

$$M(t) = \sum_{k=0}^{1} e^{tk} P(X=k) = 1-p+pe^{t}$$

Uniform MGF

$U \sim \text{Unif}(a,b)$, continuous r.v.

$$M(t) = E(e^{tU}) = \int_a^b \frac{1}{b-a} \cdot e^{tu}\, du$$

$$= \frac{e^{tb} - e^{ta}}{t(b-a)}$$

Why MGF is Important

- The MGF encodes the moments of an r.v.
- The MGF of an r.v. determines its distribution, like the CDF and PMF/PDF.
- MGFs make it easy to find the distribution of a sum of independent r.v.s.

Moments via Derivatives of the MGF

Theorem

Given the MGF of X, we can get the n^{th} moment of X by evaluating the n^{th} derivative of the MGF at 0: $E(X^n) = M^{(n)}(0)$.

$$M(t) = E[e^{tx}] = \int_{-\infty}^{+\infty} f(x) e^{tx} dx$$

$$M^{(n)}(t) = \int_{-\infty}^{+\infty} f(x) x^n e^{tx} dx \Rightarrow M^{(n)}(0) = \int_{-\infty}^{+\infty} x^n f(x) dx$$

$$= E(X^n)$$

MGF Determines the Distribution

Theorem

The MGF of a random variable determines its distribution: if two r.v.s have the same MGF, they must have the same distribution. In fact, if there is even a tiny interval $(-a, a)$ containing 0 on which the MGFs are equal, then the r.v.s must have the same distribution.

MGF of A Sum of Independent R.V.s

$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX} \cdot e^{tY}]$

$E[XY] = E(X)E(Y)$ $\quad = E[e^{tX}] \cdot E[e^{tY}]$

$= M_X(t) \cdot M_Y(t).$

Theorem

If X and Y are independent, then the MGF of $X+Y$ is the product of the individual MGFs:

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

MGF for Binomial & Negative Binomial

① $X \sim Bin(n,p)$, $X = X_1 + \cdots + X_n$, $X_i \overset{i.i.d.}{\sim} Bern(p)$. $1 \le i \le n$.

$M_X(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdots M_{X_n}(t) = [M_{X_1}(t)]^n = (1-p+pe^t)^n$

② $Y \sim NBin(r,p)$ $Y = Y_1 + Y_2 + \cdots + Y_r$. $Y_j \sim$ i.i.d. $Geom(p)$. $1 \le j \le r$.

$NBin(r,p)$: r个i.i.d的 $Geom(p)$

$M_Y(t) = M_{Y_1}(t) \cdot M_{Y_2}(t) \cdots M_{Y_r}(t)$

$= [M_{Y_1}(t)]^r = \left[\frac{p}{1-(1-p)e^t}\right]^r$, $\quad (1-p)e^t < 1$.

MGF of Location-scale Transformation

Theorem

If X has MGF $M(t)$, then the MGF of $a + bX$ is

$$E(e^{t(a+bX)}) = e^{at}E(e^{btX}) = e^{at}M(bt).$$

$$E(e^{t(a+bx)}) = \int_{-\infty}^{+\infty} f(x)e^{t(a+bx)}\,dx = e^{at}\int_{-\infty}^{+\infty} f(x)e^{btx}\,dx$$

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

MGF for Normal

① $Z \sim N(0,1)$. $M_Z(t) = E[e^{tZ}] = \int_{-\infty}^{+\infty} e^{tz} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz$

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}z^2} \frac{1}{\sqrt{2\pi}}\, dz = 1.$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2 + \frac{1}{2}t^2}\, dz$$

$$= e^{\frac{1}{2}t^2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2}\, dz.$$

$$= e^{\frac{1}{2}t^2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

$$= e^{\frac{1}{2}t^2}$$

$Y \sim N(\mu, \sigma^2)$

$Y = \mu + \sigma Z$, $Z \sim N(0,1)$.

$$M_Y(t) = e^{\mu t} M_Z(\sigma t)$$

$$= e^{\mu t} e^{\frac{1}{2}\sigma^2 t^2} = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Sum of Independent Poisson

$X \sim \text{Pois}(\lambda)$, $Y \sim \text{Pois}(\mu)$, X, Y independent.

$X+Y \sim \text{Pois}(\lambda+\mu)$

$X \sim \text{Pois}(\lambda)$ $P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$

$$M_X(t) = E[e^{tX}] = e^{\lambda(e^t-1)}$$

$$M_Y(t) = E[e^{tY}] = e^{\mu(e^t-1)}$$

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{(\lambda+\mu)(e^t-1)}$$

$$X+Y \sim \text{Pois}(\lambda+\mu)$$

Sum of Independent Normals

$X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$. X_1, X_2 independent

? $X_1 + X_2 \sim$

$M_{X_1}(t) = e^{\mu_1 t + \frac{1}{2}\sigma_1^2 t^2}$, $M_{X_2}(t) = e^{\mu_2 t + \frac{1}{2}\sigma_2^2 t^2}$

$M_{X_1+X_2}(t) = M_{X_1}(t) \cdot M_{X_2}(t) = e^{(\mu_1+\mu_2)t + \frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}$

$$\boxed{X_1 + X_2 \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)}$$

Sum is Normal

Under the setting $X_1, X_2 \sim$ i.i.d. with MGF $M_X(t)$.

and $X_1 + X_2 \sim N(0,1)$

$$M_{X_1+X_2}(t) = e^{\frac{1}{2}t^2} = M_{X_1}(t) M_{X_2}(t) = [M_X(t)]^2$$

$$\Rightarrow M_X(t) = e^{\frac{1}{4}t^2} \quad \forall t, \quad X_1, X_2 \sim N(0, \tfrac{1}{2}).$$

X_1 and X_2 are independent, $X_1 + X_2$ is normal.

$\Rightarrow$ X_1 and X_2 are normals.

Outline

Generating Functions

- Three kinds of generating functions
  - Probability Generating Functions (PGF): related to Z-transform
  - Moment Generating Function (MGF): related to Laplace transform
  - Characteristic Functions (CF): related to Fourier transform

Recall: Probability Generating Function

(Z-transform)

discrete

Definition

The *probability generating function* (PGF) of a non-negative integer-valued r.v. X with PMF $p_k = P(X = k)$ is the generating function of the PMF. By LOTUS, this is

$$E(t^X) = \sum_{k=0}^{\infty} p_k t^k.$$

The PGF converges to a value in [-1,1] for all t in $[-1, 1]$ since $\sum_{k=0}^{\infty} p_k = 1$ and $|p_k t^k| \leq p_k$ for $|t| \leq 1$.

Motivation of Characteristic Function

处理离散 ⇒ 一定在[-1,1]收敛

- **Probability generating functions(PGF)**: handling non-negative integral random variables
- **Moment generating functions(MGF)**: handling random variables 离/连均可
- Some integrals of MGF may not be finite
- **Characteristic Function**: equally useful with MGF and guarantee finiteness

Characteristic Function

Definition

The characteristic function of a random variable X is the function $\phi : \mathbb{R} \to \mathbb{C}$ defined by

$$\phi(t) = E(e^{itX}), i = \sqrt{-1}.$$

Applications of Generating Functions

- An easy way of calculating the moments of a distribution.
- Powerful tools for addressing certain counting and combinatorial problems.
- An easy way of characterizing the distribution of the sum of independent random variables.
- Tools for dealing with the distribution of the sum of a random number of independent random variables.

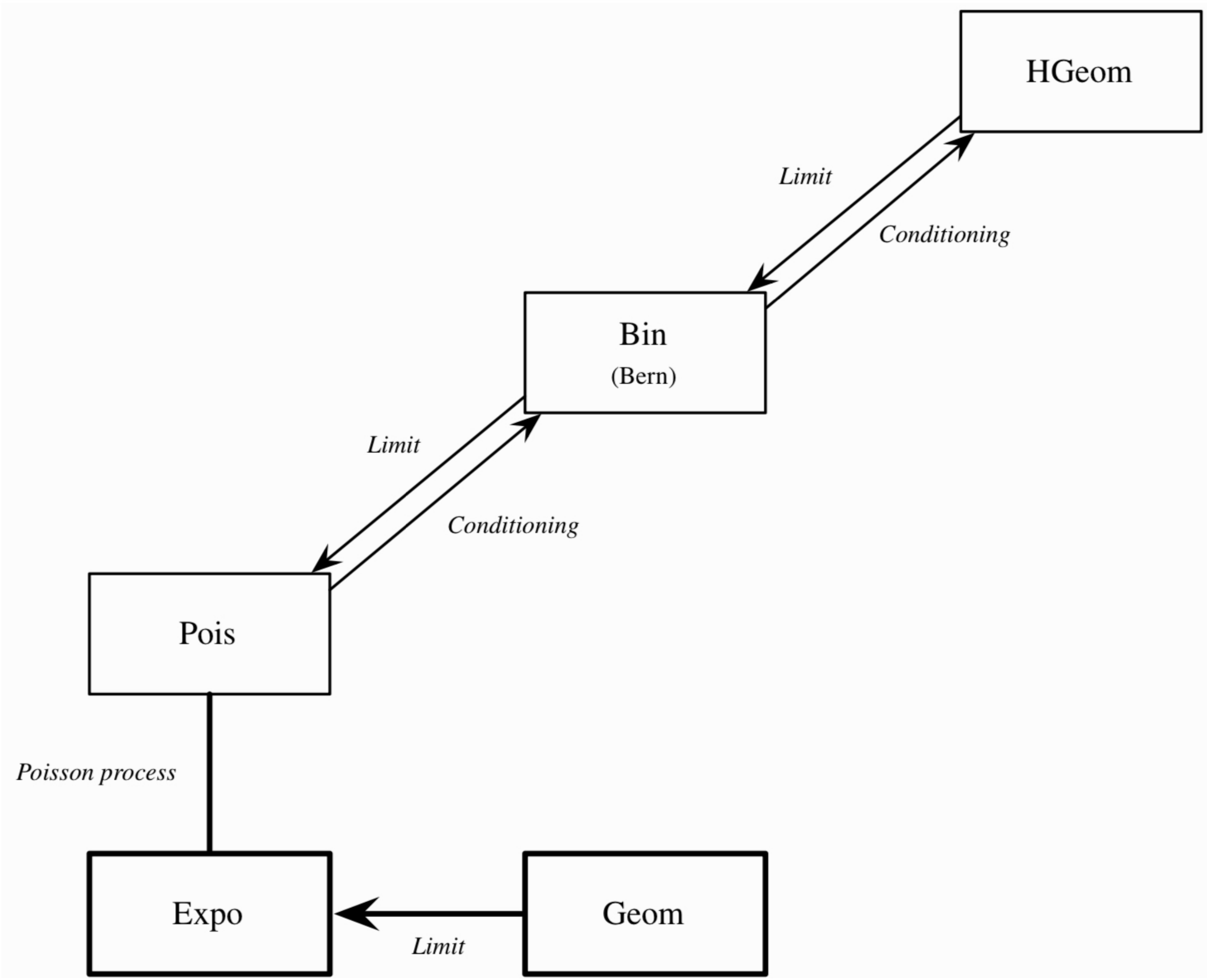
Applications of Generating Functions

- Play a central role in the study of branching processes
- Provide a bridge between complex analysis and probability
- Play a key role in large deviations theory, that is, in studying the asymptotic of tail probabilities of the form $P(X \geq c)$, when c is a large number
- Powerful tools for proving limit theorems, such as laws of large numbers and the central limit theorem

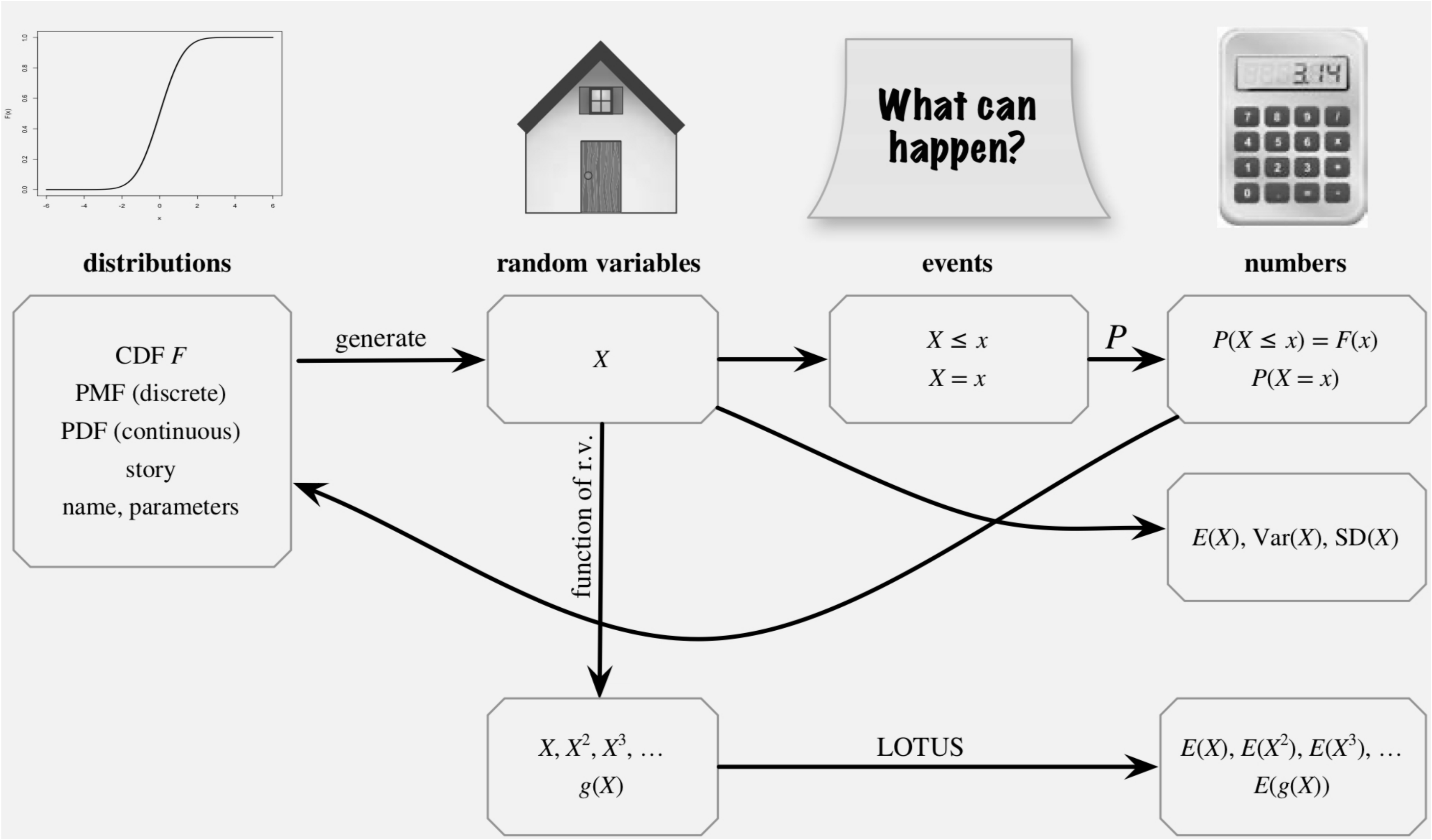
Summary 1

	Discrete r.v.	Continuous r.v.
CDF	$F(x) = P(X \leq x)$	$F(x) = P(X \leq x)$
PMF/PDF	$P(X = x)$ is height of jump of F at x • PMF is nonnegative and sums to 1: $\sum_x P(X = x) = 1$. • To get probability of X being in some set, sum PMF over that set.	$f(x) = \frac{dF(x)}{dx}$ • PDF is nonnegative and integrates to 1: $\int_{-\infty}^{\infty} f(x)dx = 1$. • To get probability of X being in some region, integrate PDF over that region.
Expectation	$E(X) = \sum_x xP(X = x)$	$E(X) = \int_{-\infty}^{\infty} xf(x)dx$
LOTUS	$E(g(X)) = \sum_x g(x)P(X = x)$	$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$

Summary 2


HGeom
Limit
Conditioning
Bin
(Bern)
Limit
Conditioning
Pois
Poisson process
Expo
Geom
Limit

Summary 3


What can happen?
distributions
random variables
events
numbers
CDF F
PMF (discrete)
PDF (continuous)
story
name, parameters
generate
X
X ≤ x
X = x
P
P(X ≤ x) = F(x)
P(X = x)
function of r.v.
E(X), Var(X), SD(X)
X, X², X³, …
g(X)
LOTUS
E(X), E(X²), E(X³), …
E(g(X))

References

- Chapters 5 & 6 of **BH**
- Chapter 3 of **BT**