Lecture 5: Continuous Random Variables

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

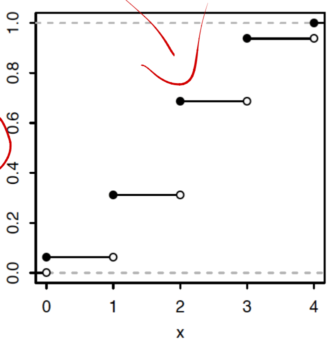
April 9, 2024

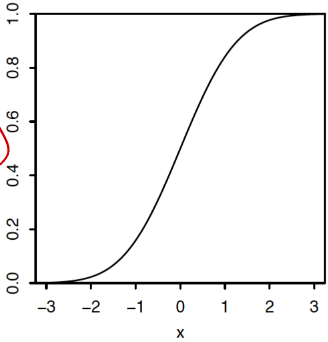
Outline

Outline

Discrete vs. Continuous

$P(X \le x)$





Continuous Random Variables

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 PMF

PDF

Definition

For a continuous r.v. X with CDF F , the *probability density function* (PDF) of X is the derivative 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 $f(x) > 0$.

Illustration of PDF

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

if $a=b$; $P(X=a) = 0$

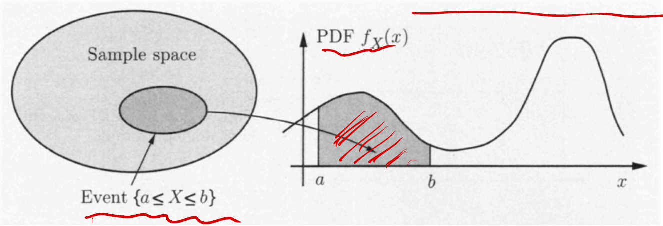
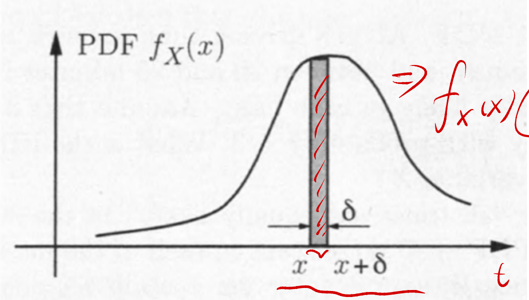


Illustration of PDF

$$P(x \le X \le x+\delta) = \int_x^{x+\delta} f_X(t)\,dt$$

$$\approx f_X(x)\cdot\delta$$



$$\Rightarrow f_X(x) \approx \frac{P(x \le X \le x+\delta)}{\delta}$$

$$\lim_{\delta \to 0} \frac{P(x \le X \le x+\delta)}{\delta} = f_X(x)$$

PDF vs. PMF

① PMF $P(X=x)$ prob. $\in [0,1]$

PDF $f(x) \neq$ prob $= \int_A f(x)dx$

$f(x) > 1$ could be.

② if X is a discrete r.v, $a \in$ support of X.

$P(X=a) > 0$.

if X is a continuous r.v.

$P(X=a) = 0, \quad \forall a \in R$.

$P(X=3.1415926) = 0$

PDF to CDF

CDF → PDF $F'(x) = f(x)$

PDF → CDF $\int_{-\infty}^{x} f(t)dt$

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 is a continuous r.v.

① $P(a<X<b) = P(a<X\le b) = P(a\le X<b) = P(a\le X\le b)$

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

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

② if X, Y are continuous r.vs. ~~$P(X=Y)=0$~~

if X, Y are independent continuous r.v.

$P(X=Y)=0$

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

Valid PDFs

Theorem

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

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

Example: Logistic Distribution

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

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

The logistic distribution has CDF

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

Find the pdf.

Example: Rayleigh Distribution

The Rayleigh distribution has CDF

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

Find the pdf.

$$f(x) = \begin{cases} F'(x) = xe^{-\frac{1}{2}x^2} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

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 $\mathbf{P}\big([x, x+\delta]\big) \approx f_X(x) \cdot \delta$.
- For any subset B of the real line,

$$\mathbf{P}(X \in B) = \int_B f_X(x)\,dx.$$

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

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

g :

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

Expectation via Survival Function

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

Theorem

Let X be a continuous and nonnegative 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

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

Proof

① $f_X(\cdot)$: PDF of X.

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

Fubini's theorem.

$X \geq 0$

$x \leq y < \infty$

$\Rightarrow$ $y \geq 0,\ y < \infty$

$0 \leq x \leq y$

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

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

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

$$= E(X)$$

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

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

② $n=3$; X_1, X_2, X_3. $3!=6$.
$P(X_1<X_2<X_3)=P(X_1<X_3<X_2)=\frac{1}{6}$.

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

Theorem

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

Key result: If X_i and X_j are independent, continuous r.v,
$P(X_i = X_j) = 0$.

Proof

Remark:

① X_1, X_2 i.i.d. r.v.s. Continuous.

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

② X_1, X_2 i.i.d. r.v.s. discrete.

$P(X_1 = X_2) \neq 0$;

$P(X_1 < X_2) = P(X_2 < X_1) = \frac{1 - P(X_1 = X_2)}{2}$.

Outline

Uniform Distribution

Definition

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

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

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

CDF of Uniform Distribution

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

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

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

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

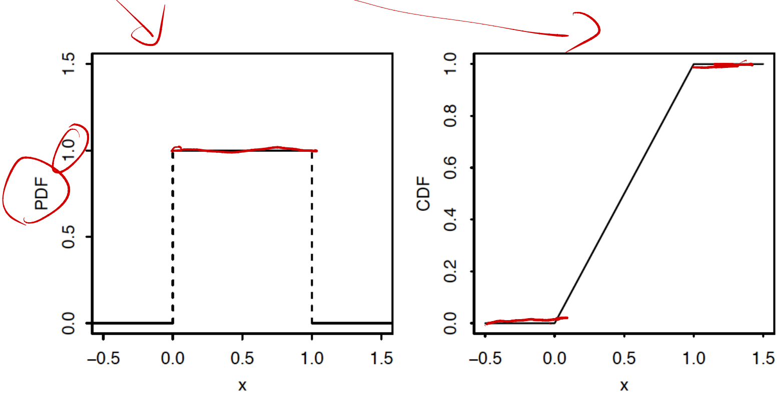
PDF $f(t) = \begin{cases} 1 & 0 < t < 1 \\ 0 & \text{otherwise} \end{cases}$

$$F_U(x) = P(U \le x) = \begin{cases} 0 & \text{if } x \le 0 \\ x & \text{if } 0 < x < 1 \\ 1 & x \ge 1 \end{cases}$$

CDF.

PDF & CDF

Unif(0,1)



Example

① Y r.v. $0 \le Y \le 1$.

② $G(y) = P(Y > y)$ $\quad$ $y \ge 1$: $G(y) = 0$.

$0 \le y \le 1$: $G(y) = P(Y > y) = P(\min(X_1, \dots X_n) > y)$

$= P(X_1 > y, X_2 > y, \dots X_n > y) = [P(X_1 > y)]^n$

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

$= [1 - P(X_1 \le y)]^n$

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

$= [1 - y]^n$.

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 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 Unif(0, 1) r.v.
- Other names:
  - probability integral transform
  - inverse transform sampling
  - the quantile transformation
  - the fundamental theorem of simulation

Universality of the Uniform

① fact: $U \sim \text{Unif}(0,1)$, $P(U \le y) = y$, $0 \le y \le 1$

$U \sim \text{Unif}(0,1)$, $X = F^{-1}(U)$.

CDF of X: $F_X(x) = P(X \le x) = P(F^{-1}(U) \le x) = P(U \le F(x))$

Theorem

$0 \le F(x) \le 1$ $\forall x$: $= F(x)$.

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 $U \sim \text{Unif}(0,1)$ 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 $F(X) \sim \text{Unif}(0,1)$.*

Proof: Universality of the Uniform

② $Y = F(X) \sim \text{unif}(0,1)$.

1°. $Y \in [0,1]$. $\quad P(Y \le y) = 0 \;;\; y \le 0$

$P(Y \le y) = 1 \;;\; y > 1$

$P(X \le x) = F(x)$;

2°. $y \in (0,1)$. $\quad P(Y \le y) = P[F(X) \le y] = P[X \le F^{-1}(y)]$

CDF of X is F

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

$= y$.

$Y \sim \text{unif}(0,1)$. $\Rightarrow F(X) \sim \text{unif}(0,1)$

Example: Universality with Logistic

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

② $F^{-1}(x) = \log \frac{x}{1-x}$, $x \in (0,1)$

③ $U \sim \text{unf}(0,1) \Rightarrow F^{-1}(U) = \log\left(\frac{U}{1-U}\right)$

$\sim$ logistic.

CDF: $P\left(\log\left(\frac{U}{1-U}\right) \le x\right) = P\left(\frac{U}{1-U} \le e^x\right) = P\left(U \le \frac{e^x}{1+e^x}\right)$

$\frac{e^x}{1+e^x} \in (0,1)$

$= \frac{e^x}{1+e^x} = F(x)$

Histogram

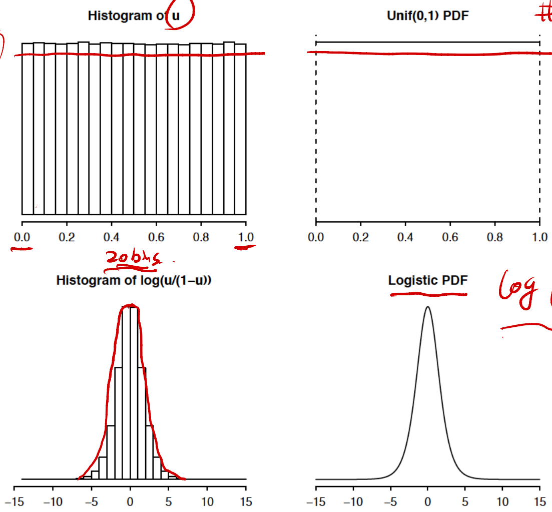
- Introduced by Karl Pearson
- A graphical representation of the distribution of numerical data
- An estimate of the probability distribution (density estimation) of a continuous variable
- To construct a histogram, the first step is to "bin" the range of values: divide the entire range of values into 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 ~ Unif(0,1)

# of samples ↑ ∞

# of bins ↑ ∞



Example: Universality with Rayleigh

① CDF of Rayleigh Distribution

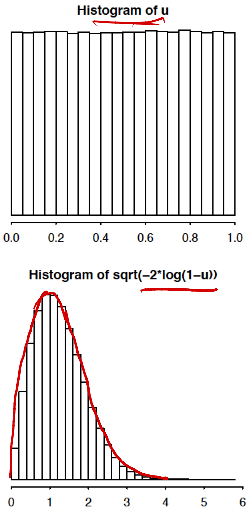
$$F(x) = 1 - e^{-\frac{1}{2}x^2}, \quad x > 0.$$

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

③ $U \sim \text{unif}(0,1)$ $\quad 1-U \sim \text{unif}(0,1)$

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

Histogram & PDF



Outline

Exponential Distribution

Definition

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

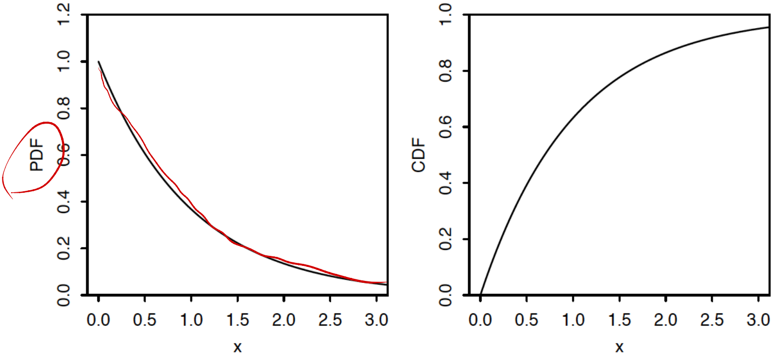
$$f(x) = \lambda e^{-\lambda x}, \ x > 0.$$

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

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

Survival function $G(x) = 1 - F(x) = e^{-\lambda x}$, $x > 0$.

Expo(1) PDF & CDF



Memoryless Property

Exponential distribution is memoryless.

$X \sim \text{Expo}(\lambda)$; $P(X \geq t) = P(X > t) = e^{-\lambda t}$, $t > 0$

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

Definition

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

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

$s = 80$, $t = 20$.

for all $s, t > 0$.

Memoryless Property

Minimum of independent Expos

① $X_j \sim \text{Expo}(\lambda_j)$; $P(X_j > t) = e^{-\lambda_j t}$, $t > 0$.

② $\forall t > 0$, $P(L > t) = P(\min(X_1, \ldots X_n) > t)$

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

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

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

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

node. Countdown clock.

1 $X_1 \sim \text{Expo}(\lambda_1)$

j $X_j \sim \text{Expo}(\lambda_j)$

n $X_n \sim \text{Expo}(\lambda_n)$

the first clock reading zero

$L = \min(X_1, \ldots X_n) \sim \text{Expo}(\lambda_1 + \cdots + \lambda_n)$

Failure (Hazard) Rate Function

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

$\frac{P(X \in (t, t+dt) \mid X > t)}{dt} = \frac{P[X \in (t, t+dt)]}{P[X > t]} \approx \frac{f(t) \cdot dt}{1 - F(t)}$

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

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

.

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

$X \sim Expo(\lambda)$, $f(t) = \lambda e^{-\lambda t}$; $F(t) = 1 - e^{-\lambda t}$

$\Rightarrow r(t) = \lambda$

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. $s, t \geq 0$

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

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

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

Theorem

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

② $G(x) = P(X > x) = P(X \geq x)$

$\Rightarrow G(s+t) = G(s) \cdot G(t), \quad \forall s, t \geq 0$

$G'(s+t) = G'(s) \cdot G(t)$.

$s \to 0$. $\Rightarrow G'(t) = G'(0) \cdot G(t)$

Proof

③ $G'(t) = G'(0) \cdot G(t)$

$= -\lambda G(t)$

$\Rightarrow \frac{dG(t)}{dt} = -\lambda G(t), \lambda > 0$

$\Rightarrow \frac{dG(t)}{G(t)} = -\lambda dt$

$\Rightarrow G(t) = e^{-\lambda t} \cdot C$

$\Rightarrow G(t) = e^{-\lambda t}$;

$G(x) = P(X > x) = 1 - F(x)$ (CDF of X)

$G'(x) = -f(x)$. PDF of X.

$G'(0) = -f(0) < 0$.

$\lambda \triangleq -G'(0) > 0$.

$G(0) = C = P(X > 0) = 1$

CDF $F(t) = 1 - e^{-\lambda t}$.

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

Geometric Distribution is also Memoryless

- 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."

Outline

Standard Normal Distribution

Definition

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

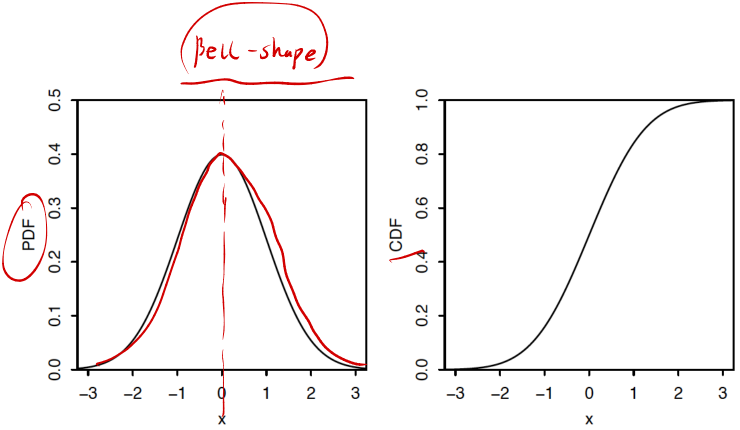
$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty.$$

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


Bell-shape
PDF
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
x
CDF
0.0
0.2
0.4
0.6
0.8
1.0
−3
−2
−1
0
1
2
3
x

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: $\Phi(z) = 1 - \Phi(-z)$.
- 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.

Verify the Validity of PDF

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

$f(z)$

① $$\left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz\right)^2 = \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)} dx\, dy = \int_0^{2\pi} \left(\int_0^{\infty} e^{-\frac{1}{2}r^2} r\, dr\right) d\theta = 2\pi.$$

② $x = r\cos\theta \qquad 0 \le \theta \le 2\pi$

$y = r\sin\theta \qquad 0 \le r < \infty$

$$dx\, dy = r\, dr\, d\theta$$

③ $$\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz = \sqrt{2\pi} \Rightarrow \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 1$$

Mean & Variance

Normal Distribution

$E(X) = \mu + \sigma E[Z] = \mu + \sigma \cdot 0 = \mu.$

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

$= \sigma^2 Var(Z)$

$= \sigma^2 \cdot 1 = \sigma^2$

Definition

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

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

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

1°. $F(x) = P(X \le x)$ $\quad Z \sim N(0,1)$

$= P(\mu + \sigma Z \le x)$

$= P\left(Z \le \frac{x-\mu}{\sigma}\right)$

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

Theorem

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

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

2°. $f(x) = F'(x)$

CDF of $N(0,1)$

and the PDF of X is

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

PDF of $N(0,1)$

68-95-99.7% Rule

$P(|X-\mu| \geq a) \leq e^{-c \cdot a^2}$

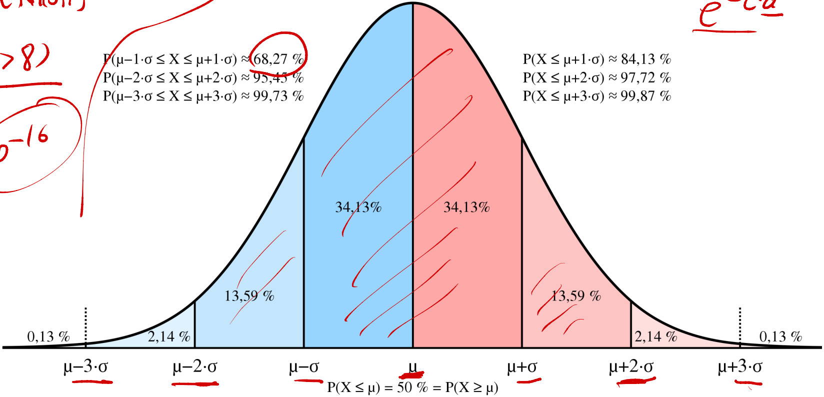
$X \sim N(\mu, \sigma^2)$

$e^{-c \cdot a}$

$X \sim N(0,1)$

$P(X > 8)$

10^{-16}



Outline

Sample Mean

$$E(\bar{X}_n) = \frac{1}{n} \cdot n E(X_1) = \mu.$$

$$Var(\bar{X}_n) = \frac{1}{n^2} Var\left(\sum_{j=1}^{n} X_j\right) = \frac{1}{n^2} Var(X_1) \cdot n = \frac{1}{n}\sigma^2$$

Definition

Let $X_1, \ldots, X_n$ 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.

$n \to \infty$ $\quad \frac{1}{n}\sigma^2 \to 0$

Central Limit Theorem

CLT

$E(\bar{X}_n) = \mu$

$Var(\bar{X}_n) = \frac{1}{n}\sigma^2$

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

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.

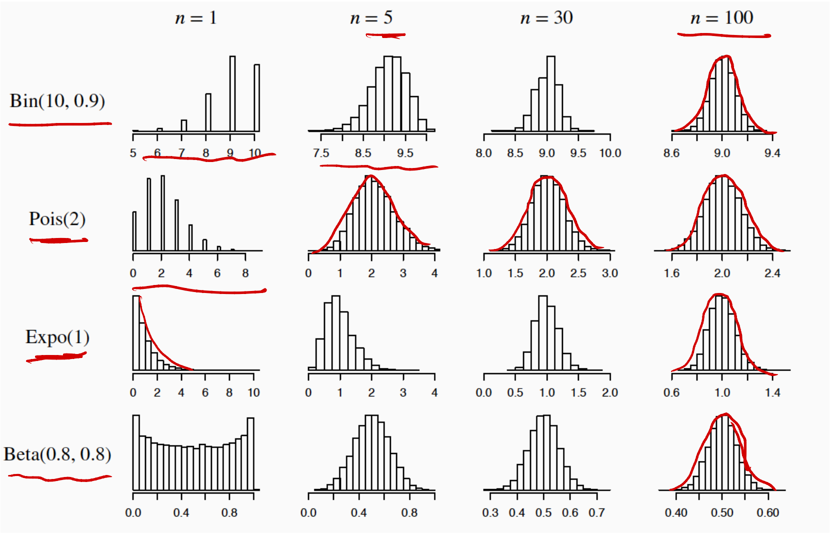
CLT Approximation

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

$X_1+X_2+\cdots+X_n \sim N(n\mu, n\sigma^2)$

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

CLT Approximation: Example



Poisson Convergence to Normal

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

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

Binomial Convergence to Normal

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

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

Continuity Correction: De Moivre-Laplace Approximation

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

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

De Moivre-Laplace Approximation

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

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

Example

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

- An exact calculation: $P(Y \leq 21) = 0.8785$
- CLT approximation:
  $P(Y \leq 21) \approx \Phi(\frac{21-np}{\sqrt{np(1-p)}}) = \Phi(1) = 0.8413$
- DML approximation:
  $P(Y \leq 21) \approx \Phi(\frac{21.5-np}{\sqrt{np(1-p)}}) = \Phi(1.17) = 0.879$

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

- Chi-Square Distribution

$$\chi_n^2 = X_1^2 + \ldots + X_n^2$$

- Student-t Distribution

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

- F-distribution:

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

Outline

Moment Generating Function

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.

Bernoulli MGF

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

$$M(t) = E[e^{tX}]$$

$$= \sum_{k=0}^{1} e^{t \cdot k} \cdot P(X=k)$$

$$= e^{t \cdot 0} \cdot P(X=0) + e^{t \cdot 1} \cdot P(X=1)$$

$$= 1-p+p \cdot e^{t}$$

Uniform MGF

$U \sim \text{unif}(a,b)$

$$M(t) = E[e^{tU}]$$

$$= \int_a^b \frac{1}{b-a} e^{tu} du$$

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

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

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

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

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 $X \sim Bin(n,p)$; $X = X_1 + \cdots + X_n$, $X_i \sim$ i.i.d. $Bern(p)$.

$$M_X(t) = M_{X_1}(t) \cdots M_{X_n}(t) = [M_{X_1}(t)]^n$$

$$= [1 - p + p \cdot e^t]^n$$

② $Y \sim NBin(r,p)$; $Y = Y_1 + \cdots + Y_r$, $Y_i \sim$ i.i.d. $Geom(p)$

$$M_{Y_1}(t) = E[e^{tY_1}] = \sum_{k=0}^{\infty} e^{tk} \cdot q^k \cdot p = p \sum_{k=0}^{\infty} (e^t \cdot q)^k \quad q = 1-p$$

$$= \frac{p}{1 - e^t q} \quad (e^t \cdot q < 1)$$

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

$q = 1-p$

MGF of Location-scale Transformation

Theorem

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

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

MGF for Normal

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

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

② $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$

$M_X(t) = e^{\mu t} \cdot M_Z(\sigma t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

Sum of Independent Poisson

$X \sim pois(\lambda)$, $Y \sim pois(\mu)$, X and Y are independent.

$X+Y \sim pois(\lambda+\mu)$

1° $X \sim pois(\lambda)$, $P(X=k) = \frac{e^{-\lambda}\cdot\lambda^k}{k!}$, $k=0,1,\ldots$

$$E[e^{tX}] = \sum_{k=0}^{\infty} e^{tk} P(X=k) = \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\lambda}\cdot\lambda^k}{k!} = e^{-\lambda}\cdot \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!}$$

$$= e^{-\lambda}\cdot e^{\lambda e^t}$$

$$= e^{\lambda(e^t-1)}$$

2°. $Y \sim pois(\mu)$, $E[e^{tY}] = e^{\mu(e^t-1)}$

3°. $M_{X+Y}(t) = M_X(t)\cdot M_Y(t) = e^{(\lambda+\mu)(e^t-1)}$

$\sim pois(\lambda+\mu)$

Sum of Independent Normals

$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, X_1 and X_2 are independent.

$X_1 + X_2 \sim ?$

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

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

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

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

Sum is Normal

Cramer's theorem: X_1 and X_2 are independent, $X_1 + X_2$ is Normal. $\Rightarrow$ X_1, X_2 are Normal.

Under the setting $X_1, X_2 \sim$ i.i.d.

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

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

$$= (M_{X_1}(t))^2$$

$$\Rightarrow M_{X_1}(t) = e^{\frac{1}{4}t^2} \sim N(0, \tfrac{1}{2})$$

$$\Rightarrow X_1, X_2 \sim N(0, \tfrac{1}{2})$$

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

Definition

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

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

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

Motivation of Characteristic Function

- **Probability generating functions(PGF)**: handling non-negative integral random variables
- **Moment generating functions(MGF)**: handling general random variables
- 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.

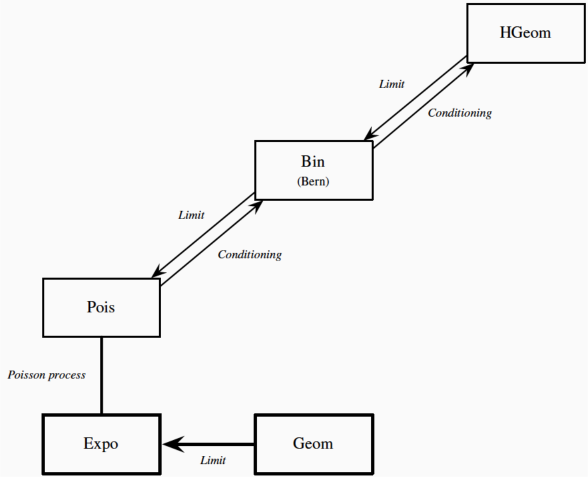
Applications of Generating Functions

- 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

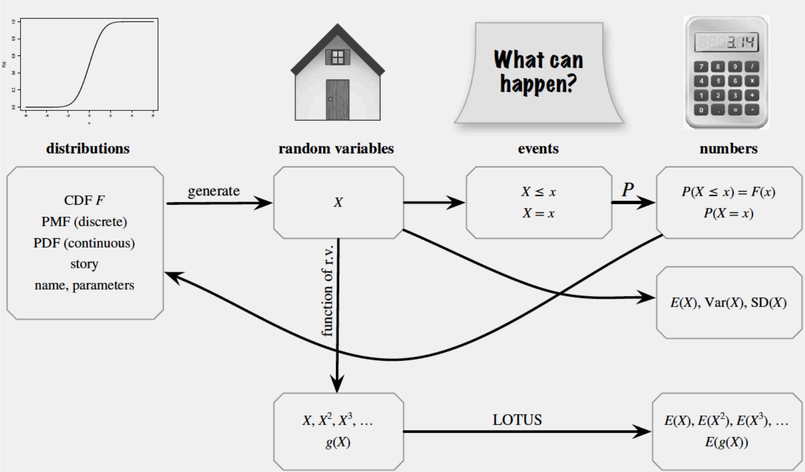
	Discrete r.v.	Continuous r.v.
CDF	$F(x) = P(X \leq x)$	$F(x) = P(X \leq x)$
PMF/PDF	$P(X = x)$ is height of jump of F at x • PMF is nonnegative and sums to 1: $\sum_x P(X = x) = 1$. • To get probability of X being in some set, sum PMF over that set.	$f(x) = \frac{dF(x)}{dx}$ • PDF is nonnegative and 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
Limit
Geom

Summary 3


What can happen?
distributions
random variables
events
numbers
CDF F
PMF (discrete)
PDF (continuous)
story
name, parameters
generate
X
$X \le x$
$X = x$
P
$P(X \le x) = F(x)$
$P(X = x)$
$E(X)$, Var(X), SD(X)
function of r.v.
$X, X^2, X^3, \ldots$
$g(X)$
LOTUS
$E(X), E(X^2), E(X^3), \ldots$
$E(g(X))$

References

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