Lecture 9: Commonly Used Distributions

Ye Tian

Department of Statistics, Columbia University
Calculus-based Introduction to Statistics (S1201)

July 19, 2022

Recap: Joint distribution and independence

- The **joint distribution** of two random variables X and Y describes how the total probability of 1 is **distributed** among all possible values of pair (X, Y). It tells us

$$\mathbb{P}(X \in \mathcal{B}_1, Y \in \mathcal{B}_2) = \mathbb{P}(\{\omega : X(\omega) \in \mathcal{B}_1\} \cap \{\omega : Y(\omega) \in \mathcal{B}_2\})$$

  for any subsets $\mathcal{B}_1$ and $\mathcal{B}_2$ of number line.
- We can define cdf, pmf and pdf for multiple r.v.'s similar to the one-variable case.
- r.v.'s X and Y are independent iff $\mathbb{P}(X \in \mathcal{B}_1, Y \in \mathcal{B}_2) = \mathbb{P}(X \in \mathcal{B}_1) \times \mathbb{P}(Y \in \mathcal{B}_2)$ for any subsets $\mathcal{B}_1$ and $\mathcal{B}_2$ of the number line, i.e.

$$\{\omega : X(\omega) \in \mathcal{B}_1\} \perp\!\!\!\perp \{\omega : Y(\omega) \in \mathcal{B}_2\}.$$

  This means independence between r.v.'s is essentially the independence between multiple pairs of events (n-"tuple" for n r.v.'s)!

Recap: Expectation and expected values

Sample mean: Data $x_1, \dots, x_n \Rightarrow \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Definition: The **expectation** of a r.v. X is defined as

$$\mathbb{E}X = \mu_X = \begin{cases} \sum_j x_j \times \mathbb{P}(X = x_j), & X \text{ is discrete,} \\ \int_{-\infty}^{+\infty} x f(x) dx, & X \text{ is continuous and } f \text{ is the pdf} \end{cases}$$

- Expectation (Or expected value) of a random variable is its **long-run average**. When $x_1, \dots, x_n \overset{i.i.d.}{\sim} X$ and $n \to +\infty$, $\bar{x} \approx \mathbb{E}X$. [1]
- $\mathbb{E}X$ indicates the "center" of X. It is not a value that is likely/expected to get.
- **Linearity of expectations:** Suppose a and b are two constants (numbers), then $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.

[1] $i.i.d.$ means "identically and independent distributed (as)"

Recap: Variance and standard deviation

Sample variance: Data $x_1, \ldots, x_n \Rightarrow s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

Definition: The **variance** of a r.v. X is defined as

$$\text{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2$$
$$= \begin{cases} \sum_j (x_j - \bar{x})^2 \times \mathbb{P}(X = x_j), & X \text{ is discrete,} \\ \int_{-\infty}^{+\infty}(x - \mathbb{E}X)^2 f(x)dx, & X \text{ is continuous and } f \text{ is the pdf} \end{cases}$$

The **standard deviation** of X is $\text{SD}(X) = \sqrt{\text{Var}(X)}$

- **A shortcut formula:** $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$
- When $x_1, \ldots, x_n \overset{i.i.d.}{\sim} X$ and $n \to +\infty$, $s^2 \approx \text{Var}(X)$.
- **"Linearity" of variances:** Suppose a and b are two constants (numbers), then:
  - $\text{Var}(aX + b) = a^2\text{Var}(X)$
  - $\text{SD}(aX + b) = |a|\text{SD}(X)$
- The variance and standard deviation is always non-negative.
- When $\text{Var}(X) = 0$, we have $\mathbb{P}(X = \mathbb{E}X) = 1$.

Recap: Covariance and correlation

Covariance between X and Y:

$$\text{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$$

The **correlation** between X and Y is defined by

$$\text{Corr}(X,Y) = \rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

- $\text{Cov}(X,Y)$ can be positive, negative, or 0, while $-1 \leq \text{Corr}(X,Y) \leq 1$
- When $\text{Cov}(X,Y) = 0$ or $\rho_{X,Y} = 0$, we say X and Y are **(linearly) uncorrelated**.
- Independence $\Rightarrow$ uncorrelation, but uncorrelation $\not\Rightarrow$ independence
- $\rho_{X,Y}$ measures the strength of **linear** association between X and Y. The larger the $|\rho_{X,Y}|$, the stronger the linear association.
- $\rho_{X,Y} > 0$ means positive association/correlation, i.e. larger X tends to associate with large Y and smaller X tends to associate with smaller Y

Recap: Population universe and sample universe

Sample universe (what we see)

Data ⇒

- Sample mean
- Sample variance
- Sample standard deviation
- Sample quantiles (sample median, quartiles...)
- Sample maximum/minimum
- Sample covariance/correlation
- Bar chart (discrete r.v.)
- Histogram (continuous r.v.)

Population universe (inaccessible)

Probability distribution ⇒

- Expectation/Mean
- Variance
- Standard deviation
- Quantiles (median, quartiles ...)
- Maximum/minimum
- Covariance/correlation
- pmf (discrete r.v.)
- pdf (continuous r.v.)

Today's goal: Know some commonly used distributions

Discrete distributions:

- Bernoulli distribution
- Binomial distribution
- Geometric distribution
- Hyper-geometric distribution
- Poisson distribution

Continuous distributions:

- Uniform distribution
- Normal distribution
- Exponential distribution
- χ^2-distribution
- t-distribution
- F-distribution

Discrete Distributions

Bernoulli distribution

Bernoulli trials: (like coin tossing)

- Each trial has two outcomes: denoted as S (success) and F (failure).
- For each trial, $\mathbb{P}(S) = p$ and $\mathbb{P}(F) = q = 1 - p$.
- The trials are independent.

Associated random variable: A r.v. X satisfies $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$, then we call X a **Bernoulli random variable** and the corresponding distribution the **Bernoulli distribution**, denoted by $X \sim \text{Bernoulli}(p)$.

- $\mathbb{E}X = p$, $\text{Var}(X) = p(1 - p)$
  Verification:
  $\mathbb{E}(X^2) = 1^2 \times p + 0^2 \times (1-p) = p \Rightarrow \text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = p - p^2$.

Examples:

- Flipping a fair coin and $X = 1$ represents the heads. $p = 1/2$.
- Quality control: Defective/Good
- Clinical trial: Survival/death
- Sampling survey: Coke/Pepsi

Bernoulli distribution

A r.v. X satisfies $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$, then we call X a **Bernoulli random variable** and the corresponding distribution the **Bernoulli distribution**.

Sometimes people call such p as the probability of success.

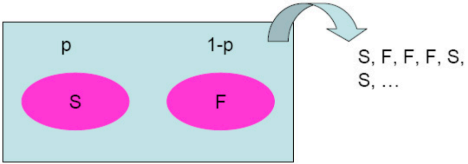
Here different values of p define different Bernoulli distributions, which we call a **family** of distributions. And we call such p as a **parameter**. Usually people first assume the data come from some distribution, then **estimate** the corresponding parameter by using the collected data.

Binomial distribution

Bernoulli trials: (like coin tossing)

- Each trial has two outcomes: denoted as S (success) and F (failure).
- For each trial, $\mathbb{P}(S) = p$ and $\mathbb{P}(F) = q = 1 - p$.
- The trials are independent.

We can carry out Bernoulli trials for many times and see how many "S" we get.



For example, flip a fair coin for 100 times and count the number of heads.

Binomial distribution

Let X be **the number of successes** in a Bernoulli process with n trials and probability of success p. Then, X follows a **Binomial distribution** and X is a **Binomial variable**, denoted by $X \sim \text{Bin}(n, p)$.

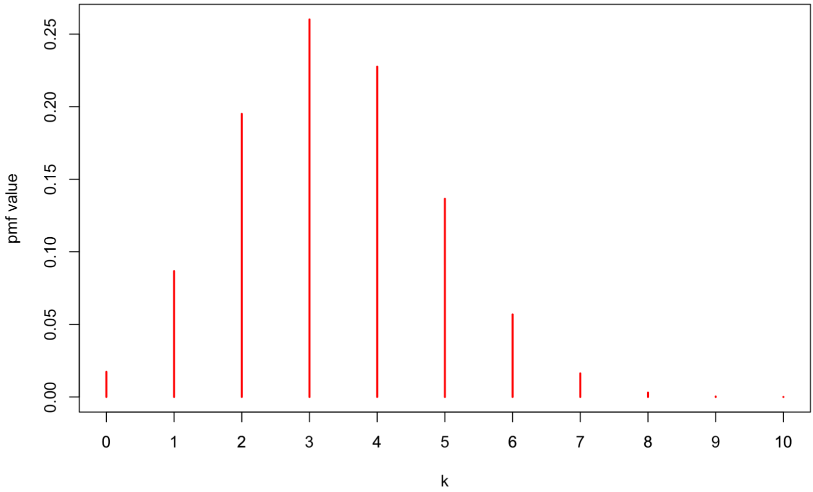
- If we denote X_i the Bernoulli r.v. corresponding to outcome of i-th Bernoulli trial, then the Binomial r.v. $X = \sum_{i=1}^{n} X_i$.
- pmf: $p(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, where $x = 0, \ldots, n$.
- cdf: $\mathbb{P}(X \leq x) = \sum_{i=0}^{x} \mathbb{P}(X = x)$
- $\mathbb{E}X = np$, $\text{Var}(X) = np(1-p)$
  Verification: $\mathbb{E}X = \mathbb{E}(\sum_{i=1}^{n} X_i) = np$.
  Since X_i's are independent, $\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(X_i) = p - p^2$.

Example:

- Flip a fair coin for $n = 100$ times and X represents the number of heads. Then $X \sim \text{Bin}(100, 1/2)$.
- Toss a die for $n = 10$ times and X represents the number of times when we get numbers 1 or 3.Then $X \sim \text{Bin}(10, 1/3)$.

Binomial distribution: pmf

pmf: $\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, where $x = 0, \ldots, n$.



pmf of Bin$(10, 1/3)$

Geometric distribution

Bernoulli trials:

- Each trial has two outcomes: denoted as S (success) and F (failure).
- For each trial, $\mathbb{P}(\mathsf{S}) = p$ and $\mathbb{P}(\mathsf{F}) = q = 1 - p$.
- The trials are independent.

We can carry out Bernoulli trials for many times and see

- Binomial distribution: how many "S" we get.
- Geometric distribution: how many trials until we first get an "S". We denote it as $X \sim \text{Geometric}(p)$.

For example:

- FSFFSF$\cdots$ $X = 2$
- FFFFFS$\cdots$ $X = 6$

pmf: $p(x) = \mathbb{P}(X = x) = \mathbb{P}(\{\underbrace{\mathsf{F}\ldots\ldots\mathsf{F}}_{(x-1)\ \mathsf{F's}}\mathsf{S}\}) = q^{x-1}p$, where $x = 1, 2, \ldots$

cdf: $F(x) = 1 - \mathbb{P}(X > x) = 1 - \mathbb{P}(\{\underbrace{\mathsf{F}\ldots\ldots\mathsf{F}}_{x\ \mathsf{F's}}\}) = 1 - q^x$

$\mathbb{E}X = 1/p$, $\text{Var}(X) = 1/p^2$

Geometric distribution

Example: A man with n keys wants to open his door and tries the keys **at random** (unsuccessful keys are not eliminated from further selections). Exactly one key will open the door.

- What's the probability that he succeeds at the 3rd time.
- What's the probability that he succeeds in 3 times.
- How many times is he expected to try until opening it?

This is a Bernoulli trial process and each time it's a success if the door is open, otherwise it's a failure.The number of trials until opening it, $X \sim \text{Geometric}(1/n)$. $(p = 1/n,\ q = 1 - 1/n)$

- $\mathbb{P}(X = 3) = (1 - 1/n)^2 (1/n)$
- $\mathbb{P}(X \leq 3) = 1 - (1 - 1/n)^3$
- $\mathbb{E}X = n$

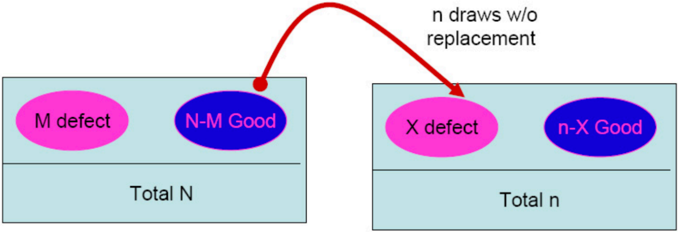
Hyper-geometric distribution

Recall: Binomial process with n trials: trials are **independent**

In a sampling survey, one typically draws **without replacement**.

Suppose there are N products in total with M defective ones and $N - M$ good ones. We sample n from these N products **without replacement**.

Let X be **the number of defective products**, which follows a **hyper-geometric distribution**.



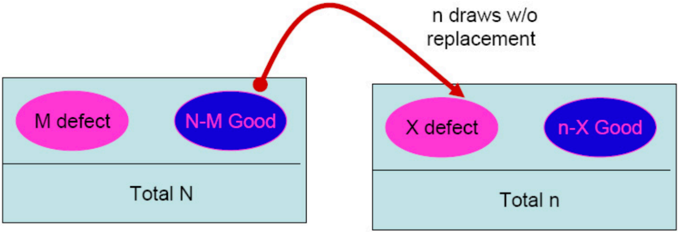
Hyper-geometric distribution

Let X be **the number of defect products**, which follows a **hyper-geometric distribution**.

pmf: $\mathbb{P}(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$, if $\max\{0, n - N + M\} \leq x \leq \min\{n, M\}$, and $\mathbb{P}(X = x) = 0$ elsewhere.

You will derive it in HW3!

$\mathbb{E}X = n \times \frac{M}{N}$.



Comparison: hyper-geometric distribution and binomial distribution

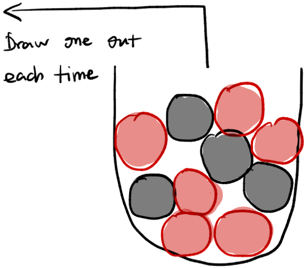
Suppose there are N products in total with M defective ones and $N - M$ good ones.

- When sampling n times with replacement, the distribution of the number of defective ones is binomial.
- When sampling n times without replacement, the distribution of the number of defective ones is hyper-geometric.
- When sampling a small fraction from a large population, the process is nearly Bernoulli trial process $\Rightarrow$ can use the binomial to approximate the hyper-geometric

Comparison: hyper-geometric distribution and binomial distribution

Example: Randomly draw 4 balls from a bag of 6 red balls and 4 black balls.

- If we draw with replacement, what's the probability of getting 2 reds ones and 2 black ones?
- If we draw without replacement, what's the probability again?
- Calculate and compare when there are 600 red balls and 400 black balls.

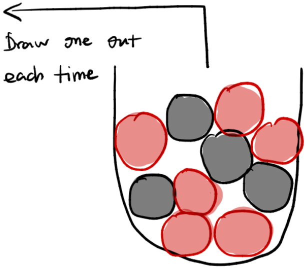


- X = number of red balls $\sim \text{Bin}(4, 0.6)$. $\mathbb{P}(X=2) = \binom{4}{2} \times 0.6^2 \times 0.4^2 = 0.3456$
- X = number of red balls $\sim$ Hypergeometric. $\mathbb{P}(X=2) = \frac{\binom{6}{2}\binom{4}{2}}{\binom{10}{4}} = 0.4285$

Comparison: hyper-geometric distribution and binomial distribution

Example: Randomly draw 4 balls from a bag where there are 6 red balls and 4 black balls.

- If we draw with replacement, what's the probability of getting 2 reds ones and 2 black ones?
- If we draw without replacement, what's the probability again?
- Calculate and compare when there are 600 red balls and 400 black balls.



- X = number of red balls $\sim \text{Bin}(4, 0.6)$.$\mathbb{P}(X = 2) = \binom{4}{2} \times 0.6^2 \times 0.4^2 = 0.3456$
- X = number of red balls $\sim$ Hypergeometric. $\mathbb{P}(X = 2) = \frac{\binom{600}{2}\binom{400}{2}}{\binom{1000}{4}} = 0.3462$
- Intuition: taking out a few balls will not affect the ball proportion much...

Poisson distribution (optional)

We say X follows **Poisson distribution** if its pmf safisfies

$$\mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots,$$

which is denoted by $X \sim \text{Poisson}(\lambda)$. And $\mathbb{E}X = \lambda$, $\text{Var}(X) = \lambda$.

Motivation: If

- The number of events occurring in any time interval is independent of the number of events in any other non-overlaping interval;
- Almost impossible for two or more events to occur simultaneously;
- The average number of occurrences per time unit is constant λ;

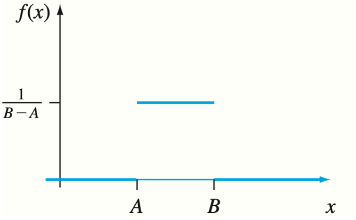
then it can be shown that the number of occurrences in a unit time interval $\sim \text{Poisson}(\lambda)$.

Continuous Distributions

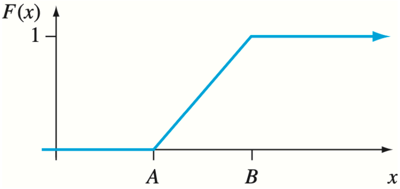
Uniform distribution

We say X follow a **uniform distribution** on $[A, B]$, if:

- Its pdf is $f(x) = \begin{cases} \frac{1}{B-A}, & A \leq x \leq B \\ 0, & \text{elsewhere} \end{cases}$
- Its cdf is $F(x) = \begin{cases} 0, & x \leq A \\ \frac{x-A}{B-A}, & A < x \leq B \\ 1, & x > B \end{cases}$



pdf



Normal distribution

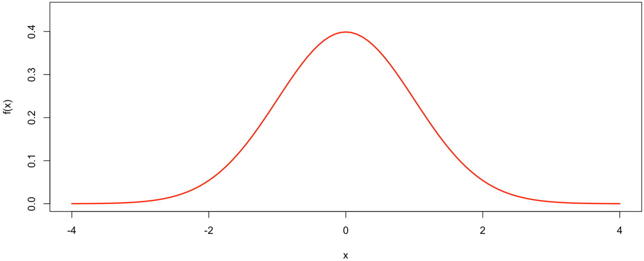
A continuous r.v. X follows a **normal distribution (or Gaussian distribution)** if its **pdf** is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < +\infty,$$

denoted by $X \sim N(\mu, \sigma^2)$. μ and σ^2 are called the mean parameter and variance parameter.

Property:

- $\mathbb{E}X = \mu$, $\text{Var}(X) = \sigma^2$.
- pdf f is symmetric around μ: $f(\mu - x) = f(\mu + x)$ for any x



pdf of $N(0, 1)$

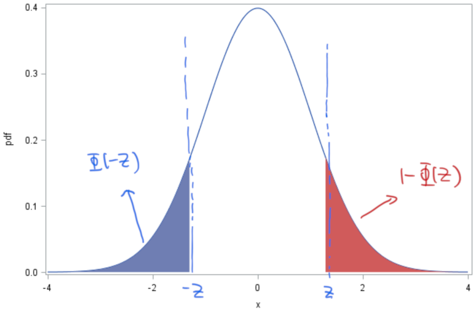
Normal distribution

$N(0,1)$ is called the **standard normal distribution** where $\mu = 0$ and $\sigma^2 = 1$. It pdf and cdf are denoted as ϕ and Φ, defined by

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \Phi(x) = \int_{-\infty}^{x} \phi(t)dt.$$

$\Phi(x)$ doesn't have an explicit expression... which is bad.

- Standardization property: If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.
- ϕ is symmetric around 0: $\phi(x) = \phi(-x)$ for any x.
- (Consequence of the second property) $\Phi(-x) = 1 - \Phi(x)$ for any x



Normal distribution

<u>**Standardization property:**</u>

- If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

<u>**Consequence:**</u>

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

Therefore, if we can calculate the cdf of the standard normal r.v. (i.e. Φ), then we can immediately calculate probability related to any normal r.v.!

The z-table

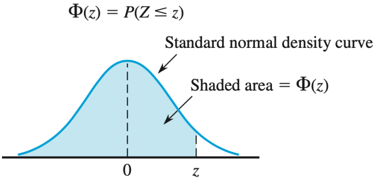
z-table lists the cdf at a fine grid of points.

Two uses:

- Look up the cdf value at a specific point;
- Given a specific cdf value, look up the corresponding quantile.

See Table A.3 of "Devore, J. L. (2011). Probability and Statistics for Engineering and the Sciences. Cengage learning. (9th edition)". You can find it on courseworks files as well.

z	.00	.01	.02	.03
−3.4	.0003	.0003	.0003	.0003
−3.3	.0005	.0005	.0005	.0004
−3.2	.0007	.0007	.0006	.0006
−3.1	.0010	.0009	.0009	.0009
−3.0	.0013	.0013	.0013	.0012



Example: normal distribution

If $X \sim N(3,4)$, compute $\mathbb{P}(1 \leq X \leq 6)$.

We know $\mu = 3$ and $\sigma = 2$ (not 4).
Step 1 (transforming X to standard normal):

$$\mathbb{P}(1 \leq X \leq 6) = \mathbb{P}\Big(\frac{1-3}{2} \leq \underbrace{\frac{X-3}{2}}_{\sim N(0,1)} \leq \frac{6-3}{2}\Big)$$
$$= \mathbb{P}(-1 \leq Z \leq 1.5)$$
$$= \Phi(1.5) - \Phi(-1)$$

Step 2 (looking up z-table): $\Phi(1.5) - \Phi(-1) = 0.9332 - 0.1587 = 0.7745.$

Example: normal distribution

If $X \sim N(3, 4)$, find its 25% quantile.

We know $\mu = 3$ and $\sigma = 2$. We want to find x that satisfies

$$\mathbb{P}(X \le x) = 0.25.$$

Step 1 (transforming X to standard normal):

$$0.25 = \mathbb{P}(X \le x) = \mathbb{P}\Big(\underbrace{\frac{X-3}{2}}_{\sim N(0,1)} \le \frac{x-3}{2}\Big)$$

$$= \mathbb{P}\Big(Z \le \frac{x-3}{2}\Big)$$

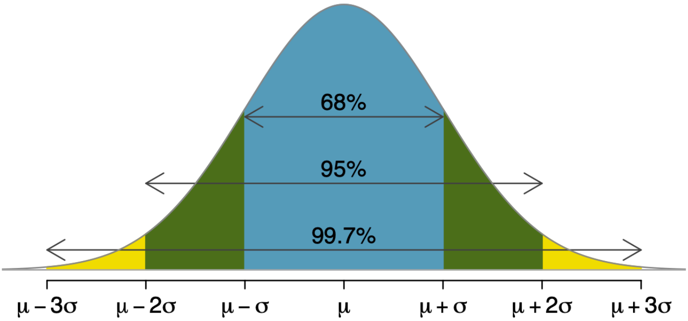
$$= \Phi\Big(\frac{x-3}{2}\Big)$$

Step 2 (looking up z-table): $\frac{x-3}{2} = \Phi^{-1}(0.25) \approx -0.67 \Rightarrow x = 1.66.$

68-95-99.7 rule (3σ-rule)

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

- $\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma) = \Phi(3) - \Phi(-3) = 0.6826 \approx 0.68$
- $\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = \Phi(2) - \Phi(-2) = 0.9544 \approx 0.95$
- $\mathbb{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = \Phi(1) - \Phi(-1) = 0.9974 \approx 0.997$



This will be useful in a wide range of practical settings, especially when trying to make a quick estimate without a calculator or z-table.

The "linearity" of expectation and variance (A general version)

Suppose $X_1, \ldots, X_n$ are random variables. $a_1, \ldots, a_n, b$ are constants.

- $\mathbb{E}[(\sum_{i=1}^n a_i X_i) + b] = \sum_{i=1}^n (a_i \mathbb{E} X_i) + b$
- If $X_1, \ldots, X_n$ are **independent**, then $\mathrm{Var}[(\sum_{i=1}^n a_i X_i) + b] = \sum_{i=1}^n a_i^2 \mathrm{Var}(X_i)$

Sum of independent normal variables

Suppose $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ and $X \perp\!\!\!\perp Y$.
Theorem: $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

Warning: This only holds when X and Y are independent!
Counter-example when $X \not\perp\!\!\!\perp Y$: $Y = -X$

Consequence:

- $X_i \sim N(\mu_i, \sigma_i^2)$. Then $\sum_{i=1}^n X_i \sim N(\sum_i^n \mu_i, \sum_{i=1}^n \sigma_i^2)$
- $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. Then $\frac{1}{n} \sum_{i=1}^n X_i \sim N(\mu, \sigma^2/n)$

Proof: By induction: $\sum_{i=1}^n X_i \sim N(\tilde{\mu}, \tilde{\sigma}^2)$. Next we will find $\tilde{\mu}$ and $\tilde{\sigma}^2$.
Since X_i's are independent: $\tilde{\mu} = \mathbb{E}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mu_i$,
$\tilde{\sigma}^2 = \text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i) = \sum_{i=1}^n \sigma_i^2$.

Review the general "linearity" of expectation and variance from last lecture!

Exponential distribution

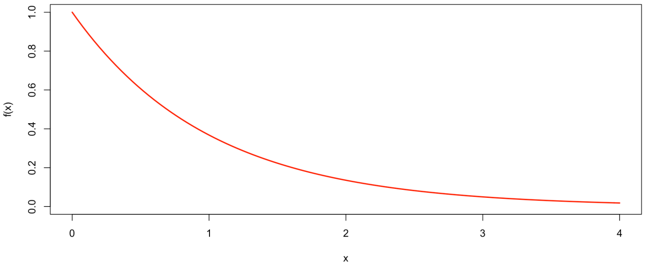
A continuous r.v. X follows a **exponential distribution** if its **pdf** is given by

$$f(x) = \lambda \exp(-\lambda x), \quad x \geq 0,$$

denoted by $X \sim \text{Exp}(\lambda)$, where λ is the parameter.

Property:

- cdf: $F(x) = \int_0^x \lambda \exp(-\lambda t) dt = 1 - \exp(-\lambda x), \quad x \geq 0.$
- $\mathbb{E}X = \int_0^x t\lambda \exp(-\lambda t) dt = \lambda^{-1}$, $\text{Var}(X) = \lambda^{-2}$. (You will prove them in HW3)



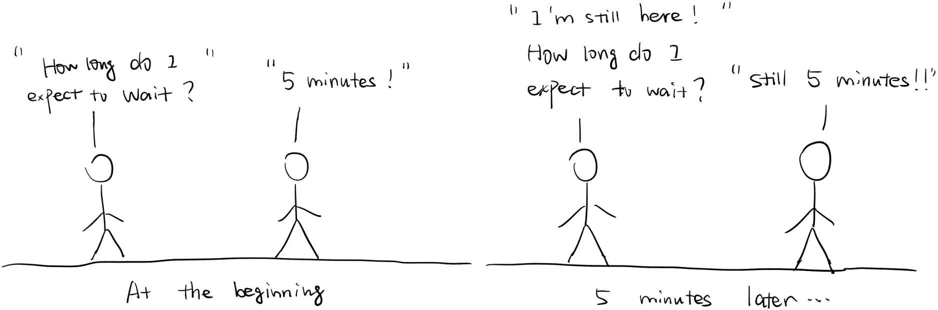
pdf of $\text{Exp}(1)$

Exponential distribution: the memoryless property

The exponential distribution has the following **memoryless property**:
Suppose $X \sim \text{Exp}(\lambda)$, then

$$\mathbb{P}(X > a + x | X > a) = \frac{\mathbb{P}(X > a + x)}{\mathbb{P}(X > a)} = \exp(-\lambda x) = \mathbb{P}(X > x),$$

for any $a \geq 0$ and $x \geq 0$.
"=": LHS $= \frac{\exp(-\lambda(a+x))}{\exp(-\lambda a)} = \exp(-\lambda x)$.



Exponential distribution

Suppose that an average of 30 customers per hour enter a store and the time between arrivals is exponentially distributed.

- Suppose the last customer just arrived. What's the probability that the next customer will show up in 3 minutes?
- Suppose that 10 minutes has passed since the last customer arrived. Will your answer change?

- On average 2 minutes elapse between successive visits. Therefore waiting time (unit: minute) $X \sim \text{Exp}(\lambda)$ and $\mathbb{E}X = \lambda^{-1} = 2 \Rightarrow \lambda = 1/2$. Thus $\mathbb{P}(X \leq 3) = 1 - e^{-3} = 0.9502$.
- Since this is an unusually long amount of time, it would seem more likely for a customer to arrive within the next minute. But by memoryless property, the waiting time from now until the next customer arrives still follows $\text{Exp}(1/2)$, therefore the answer is the same!

Many thanks to

- Yang Feng
- Joyce Robbins
- Chengliang Tang
- Owen Ward
- Wenda Zhou
- And all my teachers in the past 25 years