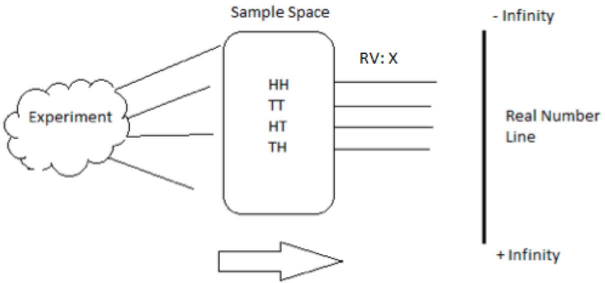
STAT 350 Lecture 4: Special Random Variables

6 Special Random Variables

(Chapter 5, 6 of WMMY)

Outline

Motivation



- Distribution of a random variable: CDF $\Longleftrightarrow$ PMF/PDF.
- Two characteristic measures for a random variable:

$$E(X), V(X).$$

Motivation

1. **Monty Hall Problem** Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?—**What's the probabilities to get a car behind each door?**
2. Flip 20 fair coins, what's the probability you get 13 Heads?
3. In a Starbucks store at campus center, according to the historical information, we know during 4-5PM on Monday, it has 30 customers on average, what's the probability it has 17 customers next Monday?

We need to learn some useful **Probability Models** people will encounter in real life!

Outline

Binomial Model

- Consider a **binary experiment** where the possible outcomes are "success" $(= 1)$ and "failure" $(= 0)$ with **probability of success** p.
- Repeat this expt for a **fixed number of** n **times** such that
  - Each repetition (trial) is **independent** of the others
  - The success probability p is the same in each expt
- Let X be the **number of successes**, then we say X is a Binomial RV with paramters n and p, and write $X \sim \text{Bin}(n, p)$.

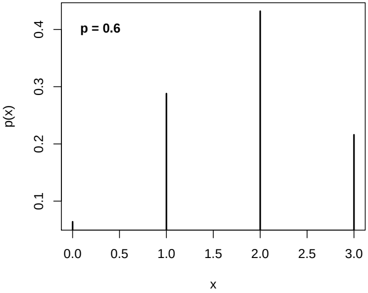
Binomial RV: number of success out of n independent trials with the same probability of success p.

Example – coin tossing

Let X denote the number of Heads in three tosses of a coin that lands on Heads with probability 0.6.

$$p(x) = \Pr(X = x) = \binom{3}{x} 0.6^x 0.4^{3-x}, \quad x = 0, 1, 2, 3.$$

x	0	1	2	3
$p(x)$	0.4^3	$3 \times 0.6 \times 0.4^2$	$3 \times 0.6^2 \times 0.4$	0.6^3



General properties of Binomial RVs

Suppose $X \sim \text{Bin}(n, p)$.

- PMF:

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

- Expected value:

$$\text{E}(X) = np.$$

- Variance:

$$\text{V}(X) = np(1-p).$$

- Calculator:
  1. $p(x) = \Pr(X = x) = \mathit{binompdf}(n, p, x),\ x = 0, 1, 2, 3, ..., n;$
  2. $F(x) = \Pr(X \le x) = \mathit{binomcdf}(n, p, x),\ x = 0, 1, 2, 3, ..., n.$

Exercise

Roll a fair die 5 times and each time when you get “1” you win 1 dollar otherwise you win nothing. Let X denote the money you win for this game.

- Write down a formula for the PMF.
- What is the probability you win exactly 1 dollar?
- What is the probability you win more than 1 dollar?
- What’s the expected money you will win?
- Close your eye and finish the game. If you are told that you won more than 1 dollar, what is the probability you win 4 dollars?

Recall

1. Flip 20 fair coins, what's the probability you get 13 Heads?
2. In a Starbucks store at campus center, according to the historical information, we know during 4-5PM on Monday, it has 30 customers on average, what's the probability it has 17 customers next Monday?

Hypergeometric Model

Consider selecting x successes from the k items labeled successes and $n - x$ failures from the $N - k$ items labeled failures when a random sample of size n is selected from N items. This is known as a **hypergeometric experiment**, that is, one that possesses the following two properties:

1. A random sample of size n is selected without replacement from N items.
2. Of the N items, k may be classified as successes and $N - k$ are classified as failures.

The number X of successes of a hypergeometric experiment is called a hypergeometric random variable, which follows the hypergeometric distribution, denoted by $X \sim \mathrm{H}(N, n, k)$.

Example – acceptance sampling

Lots of 40 components each are deemed unacceptable if they contain 3 or more defectives. The procedure for sampling a lot is to select 5 components at random and to reject the lot if a defective is found. What is the probability that exactly 1 defective is found in the sample if there are 3 defectives in the entire lot?

General properties of Hypergeometric RVs

Suppose $X \sim \mathrm{H}(N, n, k)$.

- PMF:

$$p(x) = \begin{cases} \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} & \text{if } \max\{0, n-(N-k)\} \le x \le \min\{n, k\} \\ 0 & \text{otherwise} \end{cases}$$

- Expected value:

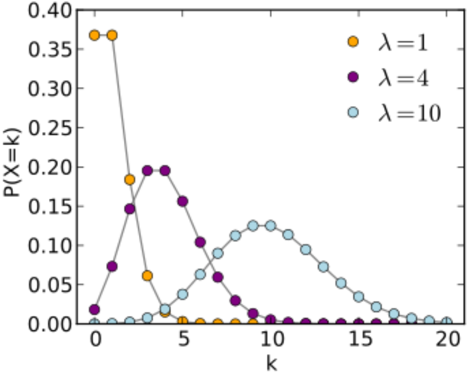
$$\mathrm{E}(X) = nk/N.$$

- Variance:

$$\mathrm{V}(X) = \frac{N-n}{N-1} \times n \times \frac{k}{N} \times (1 - \frac{k}{N}).$$

Poisson definition

We are modeling the number of occurrence of some event in a fixed time period.



Poisson RV: number of occurrence of some event in a fixed time period.

General properties of Poisson RVs

Suppose $X \sim \text{Poi}(\lambda)$ for $\lambda > 0$.

- PMF:
$$p(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$
- Expected value:
$$\text{E}(X) = \lambda.$$
- Variance:
$$\text{V}(X) = \lambda.$$
- Calculator:
  1. $p(x) = \Pr(X = x) = \mathit{poissonpdf}(\lambda, x), x = 0, 1, 2, \ldots;$
  2. $F(x) = \Pr(X \leq x) = \mathit{poissoncdf}(\lambda, x), x = 0, 1, 2, \ldots.$

Exercise

A barber shop averages 6.5 customers per hour between 12:00–5:00pm. Let X denote the number of customers arriving between 2:00–3:30pm.

1. Find the probability that exactly 6 customers arrive between 2:00 and 3:30pm.
2. Find $\Pr(X = 6 \mid X \leq 6)$.
3. Suppose the number of customers arriving between 2:00–3:30pm is recorded over a seven-day period. Assume the number of customers on different days are independent. Let Y denote the number of days with exactly 6 customers. Find $\Pr(Y \geq 1)$.

Outline

Uniform RVs

- For continuous RVs, the idea that all outcomes are equally likely is formalized by assuming the PDF is constant.
- A RV X whose PDF is constant over the interval $[a, b]$ is called a *Uniform* RV
- Notation: $X \sim \text{Unif}(a, b)$
- PDF: $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$.
- CDF: $F(x) = \frac{x-a}{b-a}$ for $a \leq x \leq b$
- Mean and variance:

$$\text{E}(X) = \frac{a+b}{2} \quad \text{and} \quad \text{V}(X) = \frac{(b-a)^2}{12}.$$

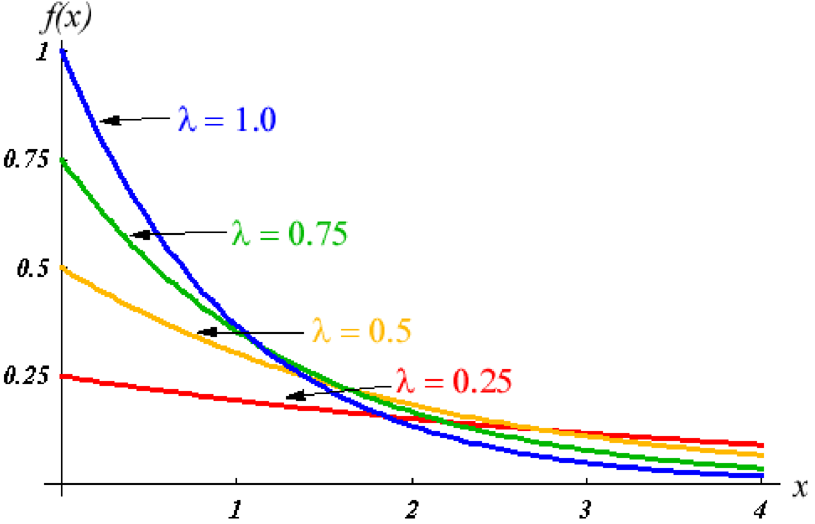
- Exercise: The random variable X is uniformly distributed on [2, 5]. Compute the value a so that $\Pr(X \geq a) = 0.7$.

Exponential RVs

- Typically used to model "waiting time" in the continuous domain.
- Examples:
  - time until next customer arrives at a bank
  - lifetime of a battery or lightbulb

Exponential RV: waiting time for some event to happen.

Exponential Distribution



Properties

- Exponential RVs have a rate parameter $\lambda > 0$.
- Notation: $X \sim \mathrm{Exp}(\lambda)$
- PDF: $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$
- CDF: $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$.
- Expected value and variance

$$\mathrm{E}(X) = \frac{1}{\lambda} \quad \text{and} \quad \mathrm{V}(X) = \frac{1}{\lambda^2}.$$

- *Lack-of-memory property.* If $X \sim \mathrm{Exp}(\lambda)$, then for any $s, t > 0$,

$$\Pr(X > s + t \mid X > t) = \Pr(X > s) = e^{-\lambda s}.$$

  - $\Pr(X > s)$: the probability you have to wait more than s mins to get the first customer
  - $\Pr(X > s + t \mid X > t)$: after waiting t mins without getting the first customer, the probability to wait more than s more mins to get the first customer

Exercises

1. Prove the lack-of-memory property.
2. Find a formula for the $100p$-th percentile of $\text{Exp}(\lambda)$.
3. A trucker hauls material between Town A and Town B. Suppose the round-trip duration X (in hours) is an Exponential RV with $\lambda = 0.05$.
   1. Find $\Pr(X > 25)$
   2. Find $\Pr(X \leq 40 \mid X > 15)$. (*hint: use lack-of-memory property*)
   3. Assume round-trip durations are independent from one trip to the next. Find the probability that exactly two of the next five round trips take more than 25 hours.

Introduction to Normal RVs

Normal RVs

- The Normal distribution is one you may have heard of in other courses—the *bell-curve*.
- It turns out to be a very reasonable approximation in many "real-life" problems, and used by researchers in all areas of science and engineering.

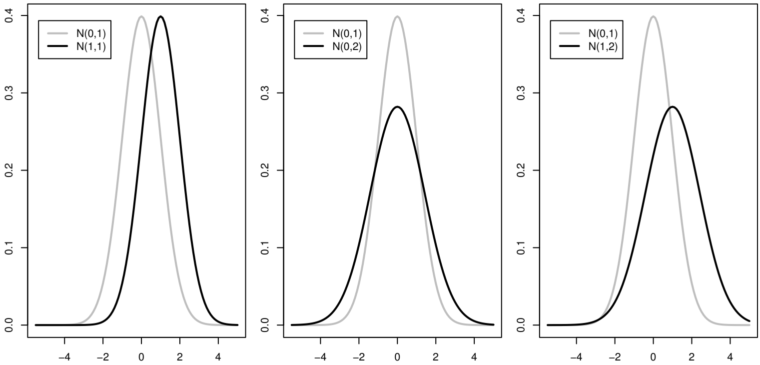
Definition

A continuous random variable X is said to have a Normal distribution with mean μ and variance $\sigma^2 > 0$ if its PDF looks like

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

We will use the notation $X \sim \mathrm{N}(\mu, \sigma^2)$.

Shapes of Normal PDFs



Linear transformations

One of the most interesting properties of the Normal distribution is that linear combinations of (not necessarily) independent Normal RVs are again Normal – more formally, the Normal distribution is *closed under linear transformations*.

Theorem

Suppose $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ *and* $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$ *are independent RVs. For constants* a*,* b*, and* c*, we have*

$$aX_1 + b \sim \mathrm{N}(a\mu_1 + b, a^2\sigma_1^2)$$
$$aX_1 + bX_2 \sim \mathrm{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

Note: Sums $X_1 + \cdots + X_n$ and means $\frac{1}{n}(X_1 + \cdots + X_n)$ are linear transformations!

Standardization

- A consequence of the above theorem is that we can *standardize* Normal RVs and retain their “Normality” property.
- This is a special case of something we have already seen – we know that the mean and variance of a standardized variable is 0 and 1, respectively.
- But if we know the variable is Normal to start with, then its standardized version is also Normal!

Corollary

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

- The special $\mathrm{N}(0, 1)$ distribution (with $\mu = 0$ and $\sigma = 1$) is the *standard Normal*.

Three Fundamental Problems in Normal Distribution

(a) Percentage: given $X \sim \mathrm{N}(\mu = 10, \sigma^2 = 9)$,

$$\Pr(4 \leq X \leq 16) = ?$$

(b) Percentile: given $X \sim \mathrm{N}(\mu = 10, \sigma^2 = 9)$, $\Pr(X \leq a) = 0.6$,

$$a?$$

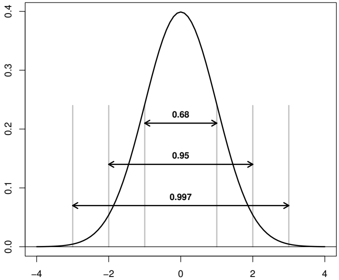
(c) Parameter: given $X \sim \mathrm{N}(\mu, \sigma^2 = 9)$, $\Pr(X \leq 4) = 0.6$,

$$\mu?$$

Recall: Rough approximations

- There are some rough approximations for the probability that X falls within k standard devations (σ's) of the mean (μ).
- The **empirical rule** or **68–95–99.7 rule** is

$$\Pr(|X - \mu| \leq k\sigma) \approx \begin{cases} 0.68 & \text{if } k = 1 \\ 0.95 & \text{if } k = 2 \\ 0.997 & \text{if } k = 3 \end{cases}$$



Three Fundamental Problems in Normal Distribution

(a) Percentage—normalcdf

When you are asked for percentage in the normal model, you will always resort to **normalcdf** in the calculator.

normalcdf($\boxed{\textbf{lower limit}}$**,** $\boxed{\textbf{upper limit}}$**,** $\boxed{\textbf{mean}}$**,** $\boxed{\textbf{standard deviation}}$**)**

$\longrightarrow$ **Percentage of the corresponding region**

(b) Percentile(Cutting-off value)-invNorm

When we are asked to find out the cutting off values or the percentiles, we always resort to **invNorm** in the calculator.

invNorm($\boxed{\textbf{Percentage on the left}}$**,** $\boxed{\textbf{mean}}$**,** $\boxed{\textbf{standard deviation}}$**)**

$\longrightarrow$ **Percentile**

(c) Parameters: μ, σ

Given some percentage and percentile information, figure out the value of parameters. Solve the following equation:

$$\frac{x-\mu}{\sigma} = invNorm(percentage, 0, 1).$$

Exercises: Percentage

1. Let $Z \sim \mathrm{N}(0, 1)$. Find the following probabilities.
   1. $\Pr(Z \leq 1.21)$
   2. $\Pr(-1.30 \leq Z \leq 0.55)$
   3. $\Pr(Z > -2.71)$
2. Suppose $X \sim \mathrm{N}(10, 9)$. Find the following probabilities.
   1. $\Pr(X > 9)$
   2. $\Pr(4 \leq X < 8)$
   3. $\Pr(X \leq 12.9)$
   4. $\Pr(X \leq -5)$

Exercises: Percentile

1. Let $Z \sim \mathrm{N}(0,1)$. Find
   1. z such that $\Pr(Z \leq z) = 0.95$
   2. z such that $\Pr(Z > z) = 0.44$
   3. z such that $\Pr(-z \leq Z \leq z) = 0.60$.
2. Given $X \sim \mathrm{N}(\mu = 10, \sigma^2 = 9)$, $\Pr(X \leq a) = 0.6$, what is a?
3. Suppose the height of adult males is Normally distributed with mean 68 inches and standard deviation 2.5 inches. How tall must a man be in order to be in the tallest 10% of adult men?

Exercises: Parameter

1. For $X \sim \mathrm{N}(\mu, \sigma^2)$, and $Z \sim \mathrm{N}(0, 1)$, verify

$$\Pr(a \leq X \leq b) = \Pr(z_a \leq Z \leq z_b),$$

   where $z_a = \frac{a-\mu}{\sigma}, z_b = \frac{b-\mu}{\sigma}$.
2. Given $X \sim \mathrm{N}(\mu, \sigma^2 = 4)$, and $\Pr(X \leq 8) = 0.7$, what is μ?
3. Given $X \sim \mathrm{N}(\mu = 10, \sigma^2)$, and $\Pr(X \leq 14) = 0.7$, what is σ^2?

Exercises: Parameter

1. Solution to (2): Due to the property in (1), $0.7 = \Pr(X \leq 8) = \Pr(Z \leq z_8)$. Thus z_8 is the 70th percentile for Z, which can be calculated by $z_8 = \text{invNorm}(0.7, 0, 1) = 0.524$. And since $z_8 = \frac{8-\mu}{\sigma} = \frac{8-\mu}{2}$, we have $\frac{8-\mu}{2} = 0.524$, which implies $\mu = 8 - 2 \times 0.524 = 6.952$.
2. Solution to (3): Due to the property in (1), $0.7 = \Pr(X \leq 14) = \Pr(Z \leq z_{14})$. Thus z_{14} is the 70th percentile for Z, which can be calculated by $z_{14} = \text{invNorm}(0.7, 0, 1) = 0.524$. And since $z_{14} = \frac{14-\mu}{\sigma} = \frac{14-10}{\sigma}$, we have $\frac{14-10}{\sigma} = 0.524$, which implies $\sigma = 4/0.524 = 7.634$ and hence $\sigma^2 = 58.272$.