Statistics and Data Analysis

Introduction to Probability (2)

Ling-Chieh Kung

Department of Information Management
National Taiwan University

Road map

- **Application: inventory management**.
- Continuous random variables.
- Normal distribution.

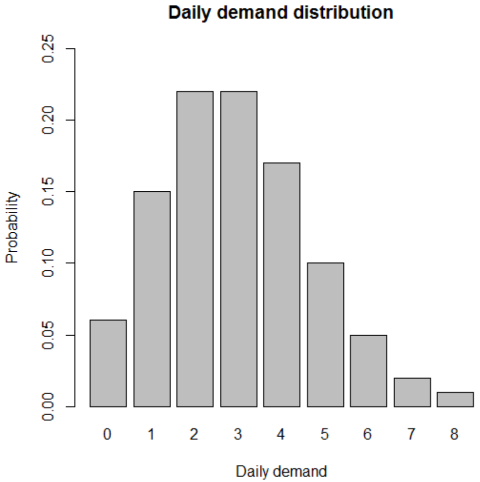
Application: inventory management

- Suppose you are selling apples.
  - The unit purchasing cost is \$2.
  - The unit selling price is \$10.
- Question: How many apples to prepare at the beginning of each day?
  - Too many is not good: **Leftovers** are valueless.
  - Too few is not good: There are **lost sales**.
- According to your historical sales records, you predict that tomorrow's demand is X, whose distribution is summarized below:

x_i	0	1	2	3	4	5	6	7	8
$\Pr(x_i)$	0.06	0.15	0.22	0.22	0.17	0.10	0.05	0.02	0.01

Daily demand distribution

- The probability distribution is depicted.
- A distribution with a long tail at the right is said to be **positively skewed**.
- It is **negatively skewed** if there is a long tail at the left.
- Otherwise, it is **symmetric**.

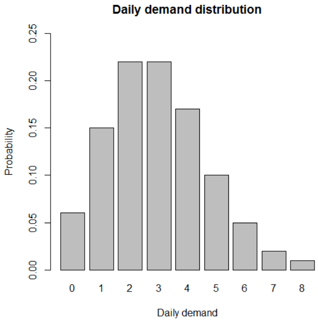


Inventory decisions

- Researchers have found efficient ways to determine the optimal (profit-maximizing) stocking level for any demand distribution.
  - This should be discussed in courses like Operations and Service Management.
- For our example, at least we may try all the possible actions.
  - Suppose the stocking level is y, $y = 0, 1, ..., 8$, what is the **expected** profit $f(y)$?
  - Then we choose the stocking level with the highest expected profit.

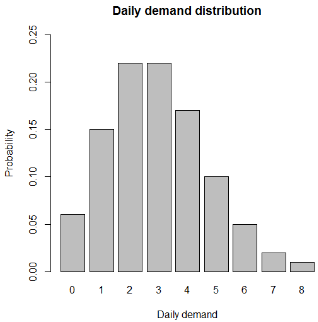
Expected profit function

- If $y = 0$, obviously $f(y) = 0$.
- If $y = 1$:
  - With probability 0.06, $X = 0$ and we lose $0 - 2 = -2$ dollars.
  - With probability 0.94, $X \geq 1$ and we earn $10 - 2 = 8$ dollars.
  - The expected profit is $(-2) \times 0.06 + 8 \times 0.94 = 7.4$ dollars.



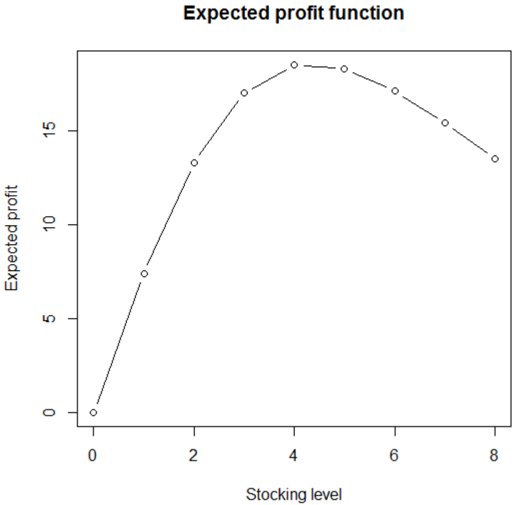
Expected profit function

- If $y = 2$:
  - With probability 0.06, $X = 0$ and we lose $0 - 4 = -4$ dollars.
  - With probability 0.15, $X = 1$ and we earn $10 - 4 = 6$ dollars.
  - With probability 0.79, $X \geq 2$ and we earn $20 - 4 = 16$ dollars.
  - The expected profit is $(-4) \times 0.06 + 6 \times 0.15 + 16 \times 0.79 = 13.3$ dollars.
- By repeating this on $y = 3, 4, ..., 8$, we may fully derive the expected profit function $f(y)$.



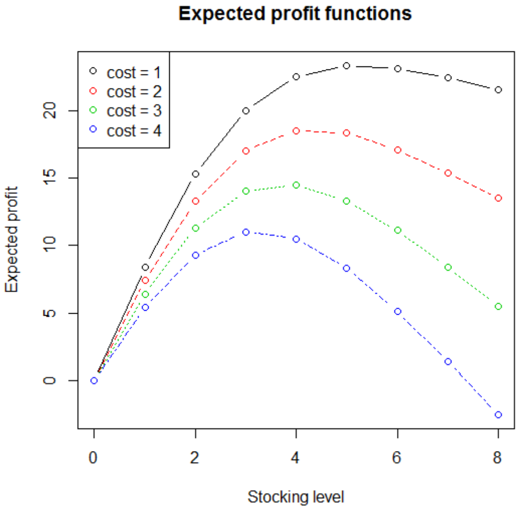
Optimizing the inventory decision

- The optimal stocking level is 4.
- What if the unit production cost is not \$2?



Impact of the unit cost

- For unit costs 1, 2, 3, or 4 dollars, the optimal stocking levels are 5, 4, 4, and 3, respectively.
- Does the optimal stocking level always decrease when the unit cost increase?
- Anyway, understanding probability allows us to make better decisions!



Road map

- Application: inventory management.
- **Continuous random variables**.
- Normal distribution.

Continuous random variables

- Some random variables are **continuous**.
  - The value of a continuous random variable is **measured**, not **counted**.
  - E.g., the number of students in our classroom when then next lecture starts is discrete.
  - E.g., the temperature of our classroom at that time is continuous.
- For a continuous RV, its possible values typically lie in an **interval**.
  - Let X be the temperature (in Celsius) of our classroom when the next lecture starts. Then $X \in [0, 50]$.
- We are interested in knowing the following quantities:
  - $\Pr(X = 20)$, $\Pr(18 \leq X \leq 22)$, $\Pr(X \geq 30)$, $\Pr(X \leq 12)$, etc.

Continuous random variables

- As another example, consider the number of courses taken by a student in this semester.
  - Let's label students in this class as 1, 2, ..., and n.
  - Let X_i be the number of courses taken by student i.
  - Obviously, X_i is discrete.
  - However, their mean $\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n}$ is (approximately) continuous!
  - Especially when n is large.
- In statistics, the understanding of continuous random variables is much more important than that of discrete ones.

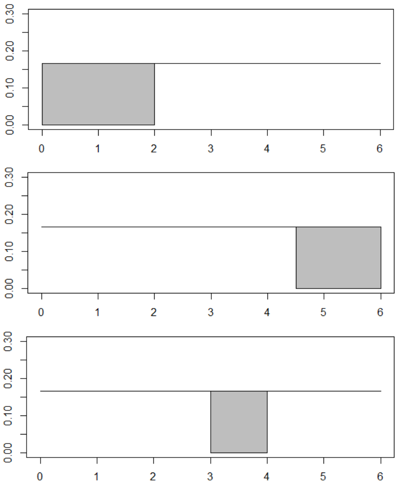
Continuous probability distribution

- Consider the task of randomly generating a value in $[0, 6]$ again.
  - Let the outcome be X.
  - All values in $[0, 6]$ are equally likely to be observed.
- What is the probability of getting $X = 2$?
  - Because all the values (0, 1, 2.4, 3.657432, 4.44..., π, $\sqrt{2}$, etc.) may be an outcome, the probability of getting **exactly** $X = 2$ is **zero**.
  - In general, $\Pr(X = a) = 0$ for all $a \in \mathbb{R}$ as long as X is continuous.
- What is the probability of getting **no greater than** 2, $\Pr(X \leq 2)$?[1]

[1] Because $\Pr(X = 2) = 0$, we have $\Pr(X \leq 2) = \Pr(X < 2)$. In other words, "less than" and "no greater than" are the same regarding probabilities.

Continuous probability distribution

- Obviously, $\Pr(X \leq 2) = \frac{1}{3}$.
- Similarly, we have:
  - $\Pr(X \leq 3) = \frac{1}{2}$.
  - $\Pr(X \geq 4.5) = \frac{1}{4}$.
  - $\Pr(3 \leq X \leq 4) = \frac{1}{6}$.
- For a continuous random variable:
  - A **single value** has no probability.
  - An **interval** has a probability!



Continuous probability distribution

- Let Y be a random variable within $[10, 50]$. Suppose that all values occur with the same likelihood.
- Then we have:
  - $\Pr(Y \leq 30) = \frac{1}{2}$.
  - $\Pr(Y \geq 45) = \frac{50-45}{50-10} = \frac{1}{8}$.
  - $\Pr(20 \leq Y \leq 30) = \frac{1}{4}$.
  - $\Pr(Y \leq 5) = 0$.
- In general, we have

$$\Pr(a \leq Y \leq b) = \frac{b-a}{50-10}$$

for all $10 \leq a \leq b \leq 50$.

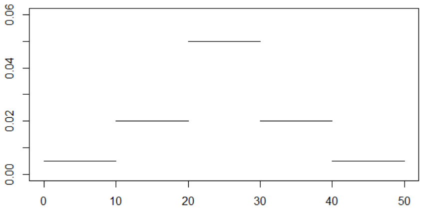
Uniform distribution

- The random variables X and Y are very special:
  - All possible values are equally likely to occur.
- For a continuous random variable of this property, we say it follows a (continuous) **uniform distribution**.
  - When X is uniformly distributed in $[a, b]$, we write $X \sim \text{Uni}(a, b)$.
  - If a discrete random variable possesses this property (e.g., rolling a fair dice), we say it follows a discrete uniform distribution.
- When do we use a uniform random variable?
  - When we want to draw one from a population fairly (i.e., randomly).
  - When we sample from a population.

Non-uniform distribution

- Sometimes a continuous random variable is not uniform.
  - Let X be the temperature of the classroom when the next lecture starts.
  - We can say that $X \in [0, 50]$.
  - X is more likely to occur in $[20, 30]$ but less likely in $[10, 20]$ and $[30, 40]$. It is almost impossible for X to be in $[0, 10]$ and $[40, 50]$.
  - The likelihood of X in different intervals can be different.
- We use a **probability density function** (pdf) $f(x)$ to describe the likelihood of each possible value. Larger $f(x)$ means **higher** likelihood.
- For X, let its pdf be

$$f(x) = \begin{cases} 0.005 & \text{if } x < 10 \\ 0.02 & \text{if } 10 \leq x < 20 \\ 0.05 & \text{if } 20 \leq x < 30 \\ 0.02 & \text{if } 30 \leq x < 40 \\ 0.005 & \text{if } 40 \leq x \end{cases}.$$

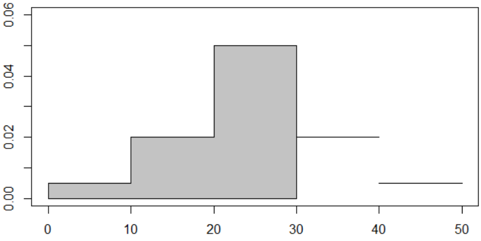


Non-uniform distribution

- Given the pdf $f(x)$, the probability for X to be smaller than any given value $x > 0$ is

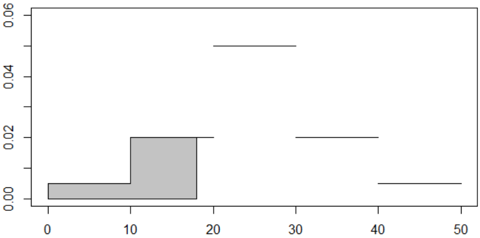
$$\Pr(X \leq x) = \int_0^x f(v)dv.$$

  That is, the **area below the pdf** from 0 to x.
  - The "sum" of the likelihood of all values between 0 to x is the probability.
- E.g., $\Pr(X \leq 30) = 10 \times 0.005 + 10 \times 0.02 + 10 \times 0.05 = 0.75$.



Non-uniform distribution

- E.g., $\Pr(X \leq 18) = 10 \times 0.005 + 8 \times 0.02 = 0.21$.



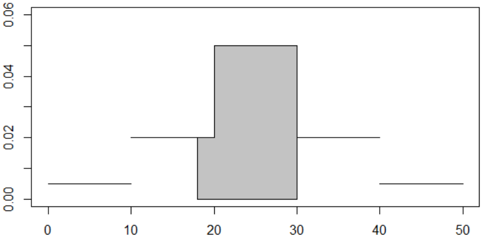
  - $\Pr(X \leq 18) = 0.21 < \frac{18}{30}\Pr(X \leq 30) = \frac{18}{30}(0.75) = 0.45$. This is because X is more likely to be within $[20, 30]$.

Cumulative distribution functions

- We call $F(x) = \Pr(X \leq x)$ the **cumulative distribution function** (cdf) of X.
  - It is the cumulative probability up to the give value x.
- For any given region $[a, b]$, we then have

$$\Pr(a \leq X \leq b) = \Pr(X \leq b) - \Pr(X \leq a) = F(b) - F(a).$$

- E.g., $\Pr(18 \leq X \leq 30) = F(30) - F(18) = 0.75 - 0.21 = 0.54$.



Density and distribution functions

- Continuous random variables are important in inferential statistics.
  - Because many statistics, e.g., the sample mean $\bar{x}$, are **continuous** random variables.
  - To infer a parameter (e.g., the population mean μ), we rely on the **distribution of a statistic**.
- The concepts of density and distribution functions (i.e., pdf and cdf) are important.
  - A pdf describes the **likelihood** of each possible value.
  - A cdf are used to calculate the **probability** of being within a region.
  - In most cases, statistical software do the calculations.
  - But we need to know **what to calculate**.

Road map

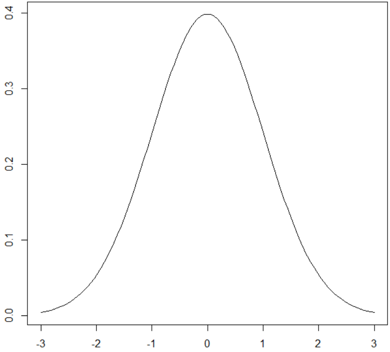
- Application: inventory management.
- Continuous random variables.
- **Normal distribution**.

Central tendency

- In practice, typically data do not spread uniformly.
- Values tend to be **close to the center**.
  - Natural variables: heights of people, weights of dogs, lengths of leaves, temperature of a city, etc.
  - Performance: number of cars crossing a bridge, sales made by salespeople, consumer demands, student grades, etc.
  - All kinds of errors: estimation errors for consumer demand, differences from a manufacturing standard, etc.
- We need a distribution with such a central tendency.

Normal distribution

- The **normal distribution** is the most important distribution in statistics (and many other fields).
  - If a random variable follows the normal distribution, most of its "normal values" will be close to the center.
- It is **symmetric** and **bell-shaped**.



Normal distribution

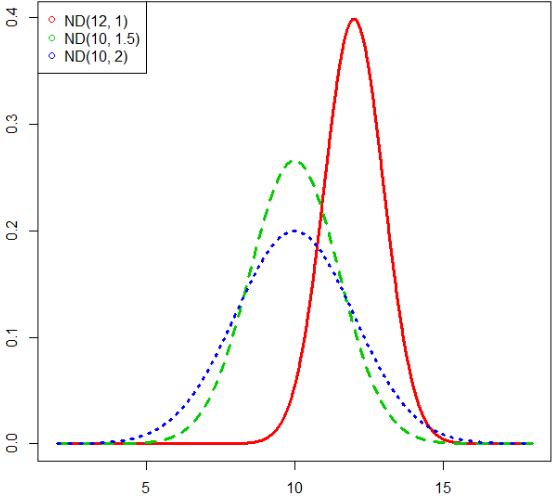
- Mathematically, a random variable X follows a normal distribution with mean μ and standard deviation σ if its pdf is

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for all } x \in (-\infty, \infty).$$

  - Well... Anyway, you know there is a definition.
  - We write $X \sim \text{ND}(\mu, \sigma)$.
- Some **important** properties of the normal distribution:
  - Its peak locates at its mean (expected value).
  - Its mean equals its median.
  - The larger the standard deviation, the flatter the curve.

Altering normal distributions

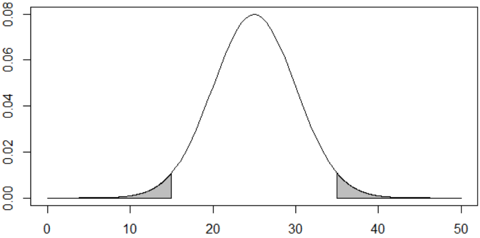
- Increasing the expected value μ shifts the curve to the right.
- Increasing the standard deviation σ makes the curve flatter.



Classroom temperature

- Let X be the room temperature when the next lecture starts.
- Suppose that $X \sim \text{ND}(25, 5)$.
- Suppose that the lecture must be canceled if $X < 15$ or $X > 35$.
- The probability for the lecture to be canceled is

$$\begin{aligned}\Pr(X < 15 \text{ or } X > 35) &= \Pr(X < 15) + \Pr(X > 35) \\ &= 2\Pr(X < 15) \approx 5\%.\end{aligned}$$



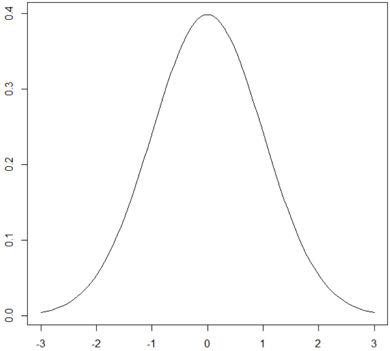
Standard normal distributions

- The **standard normal distribution**, sometimes denoted as $\phi(x)$, is a normal distribution with $\mu = 0$ and $\sigma = 1$.
- All normal distributions can be transformed to the standard normal distribution.

Proposition 1

If $X \sim \text{ND}(\mu, \sigma)$, *then* $Z = \frac{X-\mu}{\sigma} \sim \text{ND}(0, 1)$.

- This transformation is called **standardization**.



Standard normal distributions

- Consider a set of data.
- For a value x, we define its **z-score** as $z = \frac{x-\mu}{\sigma}$.
  - It measures how far this value is from the mean, using the standard deviation as the unit of measurement.
  - E.g., if $z = 2$, the value is 2 standard deviations above the mean.
- Is two σs away from the mean **normal or not**?
- Recall our classroom temperature example:
  - $X \sim \text{ND}(25, 5)$ and $\Pr(X < 15) + \Pr(X > 35) \approx 5\%$.
  - For a normally distributed random variable, it will be two σs away from mean with probability 5%.
- Recall our "3σ" rule for **detecting outliers**.
  - We also have $\Pr(X < 10) + \Pr(X > 40) \approx 0.25\%$.
  - That is why the distance of three σs is suggested.

Quality control and "six sigma"

- As long as the distribution is normal:

Quality standard	Probability
One σ	68%
Two σ	95%
Three σ	99.7%
Six σ	99.9997%

- The calculations of these probabilities depends on cdf.