GMBA 7098: Statistics and Data Analysis (Fall 2014)

Introduction to Probability (2)

Ling-Chieh Kung

Department of Information Management
National Taiwan University

October 20, 2014

Road map

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

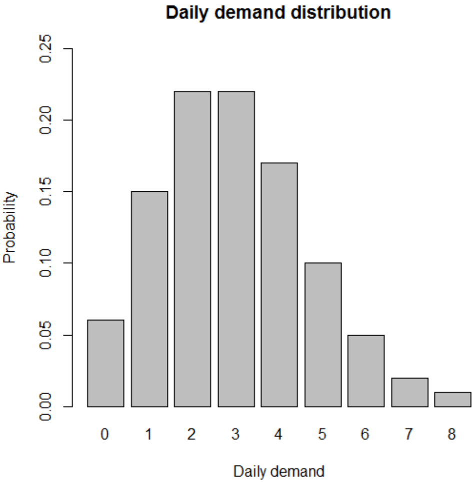
Application: inventory management

- 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**.

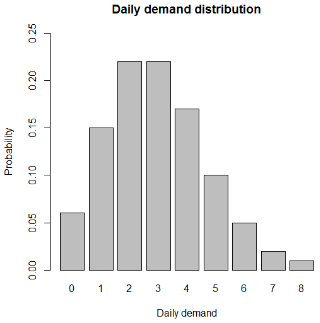


Inventory decisions

- 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.

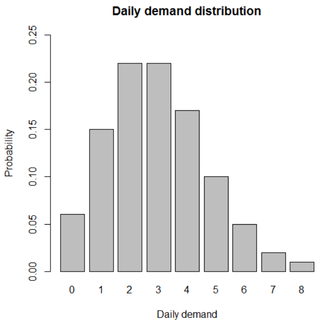
Expected profit function

- 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.



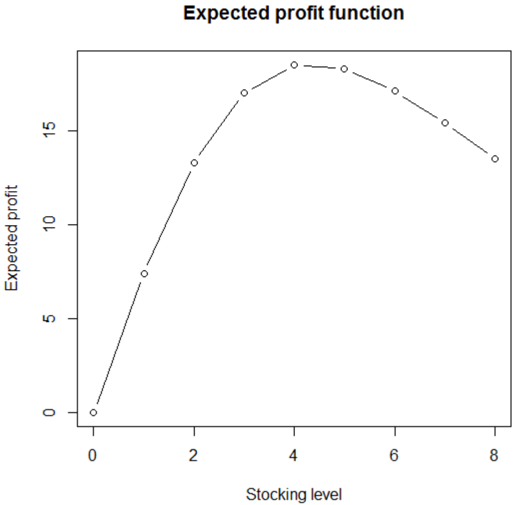
Expected profit function

- 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 $f(y)$.



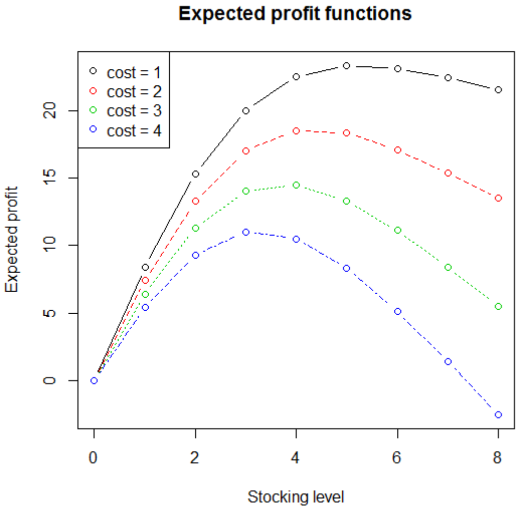
Optimizing the inventory decision

- 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

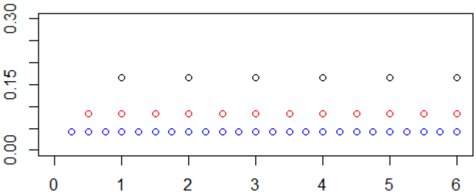
- Some random variables are **continuous**.
  - The value of a continuous random variable is **measured**, not **counted**.
  - E.g., the number of students in our classroom when then next lecture starts is discrete.
  - E.g., the temperature of our classroom at that time is continuous.
- For a continuous 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!
- In statistics, the understanding of continuous random variables is much more important than that of discrete ones.

Rolling a multi-face dice

- Let's start by, again, rolling a dice.
- Let X_1 be the outcome of rolling a fair "normal" dice, then we have $\Pr(X_1 = x) = \frac{1}{6}$ for $x = 1, 2, ..., 6$.
- Let X_2 be the outcome of rolling a fair 12-face dice with sample space $S_2 = \{\frac{1}{2}, 1, ..., \frac{11}{2}, 6\}$, then we have $\Pr(X_2 = x) = \frac{1}{12}$ for $x \in S_2$.
- Let X_3 be the outcome of rolling a fair 24-face dice with sample space $S_3 = \{\frac{1}{4}, \frac{1}{2}, ..., \frac{23}{4}, 6\}$, then we have $\Pr(X_3 = x) = \frac{1}{24}$ for $x \in S_3$.



Rolling a multi-face dice

- Let X_4 be the outcome of rolling a fair n-face dice, then we have $\Pr(X_4 = x) = \frac{1}{n}$ for $x \in S_4 = \{\frac{6}{n}, \frac{12}{n}, ..., 6\}$.
- When n approaches infinity, we may get any value within 0 and 6. However, the **probability of getting each value** is 0.
- There are **infinitely many** possible values, but the total probability is 1. Therefore, the probability of getting each value can only be 0.
- In general, for any continuous random variable X, we have

$$\Pr(X = x) = 0$$

for all x!

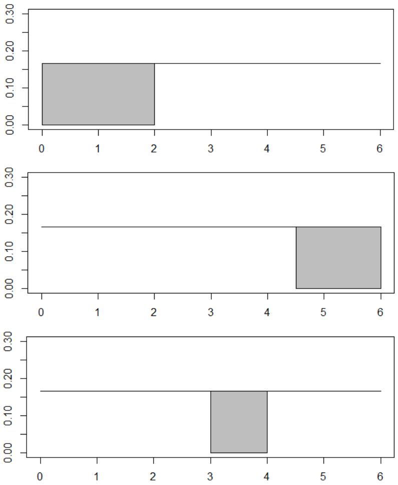
Continuous probability distribution

- Consider the example 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.
- We know the probability of getting **exactly** 2 is 0; $\Pr(X = 2) = 0$.
- 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!
- We need a formal way to describe a continuous distribution.



Probability density functions

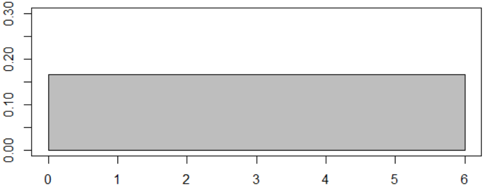
- A continuous distribution is described by a **probability density functions** (pdf).
  - A pdf is typically denoted by $f(x)$, where x is a possible value.
  - For each possible value x, the function gives the probability **density**. It is **not** a probability!
- What is that "density" for?
  - For a discrete distribution, we define a probability mass function.
  - Accumulating density gives us mass; accumulating probability density gives us probability.
- For any continuous random variable $X \in [a, b]$, its pdf $f(x)$ satisfies

$$\int_a^b f(x)dx = 1,$$

  i.e., the area under $f(\cdot)$ within $[a, b]$ must be 1.

Probability density functions

- Let X be the outcome of randomly generating a value in $[0, 6]$.
  - All values in $[0, 6]$ are equally likely to be observed.
  - They all have the same probability density: $f(x) = y$ for all $x \in [0, 6]$.
  - What is the value of y?
- The area under $f(\cdot)$ within $[0, 6]$ must be 1:



We need $6y = 1$, i.e., $f(x) = y = \frac{1}{6} \approx 0.167$.

Uniform distribution

- The random variable X is 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**.
  - 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.

Road map

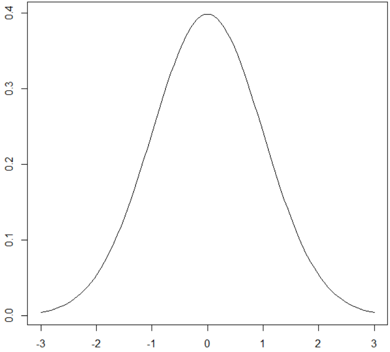
- Application: inventory management.
- Continuous random variables.
- **Normal distribution**.

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 "normal data" will be close to the center.
- It is **symmetric** and **bell-shaped**.



Normal distribution

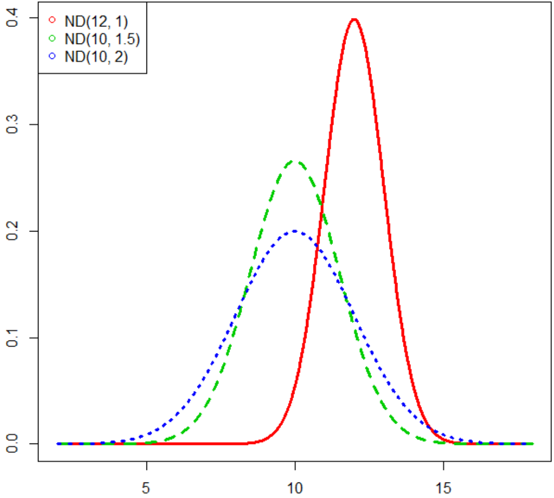
- Mathematically, a random variable X follows a normal distribution with mean μ and standard deviation σ if its pdf is

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

  - 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.



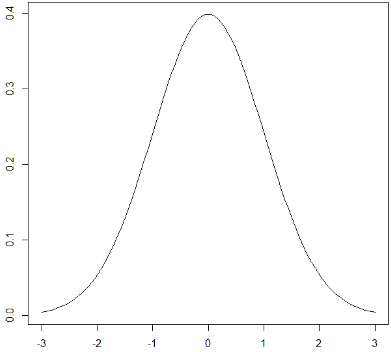
Standard normal distributions

- 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**.



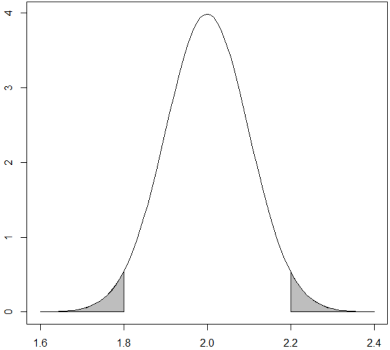
Standard normal distributions

- 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.
  - A z-score may be positive or negative.
- Is two σ away from the mean normal or not?

Quality control

- A seller sells candies in bags. She asks her son to put candies in bags and make each bag weigh 2 kg. No bag can weigh more than 2.2 kg or less than 1.8 kg.
  - Her son, unfortunately, is careless.
  - If X is the weight of a randomly drawn bag, $X \sim \text{ND}(2, 0.1)$.
  - A bag that weighs 2.2 kg is two σ above the mean.
- The probability for a bag to be "bad" is

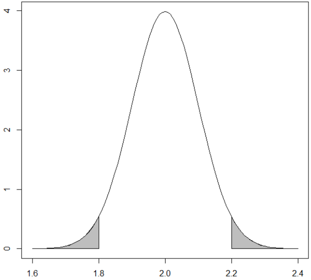
$$\begin{aligned} &\Pr(X \geq 2.2 \text{ or } X \leq 1.8) \\ =&\Pr(X \geq 2.2) + \Pr(X \leq 1.8). \end{aligned}$$



Quality control

- R helps us do the calculation.
  - `pnorm(q, mean, sd)` finds $\Pr(X \leq \mathtt{q})$ for $X \sim \text{ND}(\mathtt{mean}, \mathtt{sd})$.
- The probability for a bag to be "bad" is

$$\begin{aligned} &\Pr(X \geq 2.2 \text{ or } X \leq 1.8) \\ = &\Pr(X \geq 2.2) + \Pr(X \leq 1.8) \\ = &\texttt{pnorm(1.8, 2, 0.1) * 2} \\ \approx &5\%. \end{aligned}$$

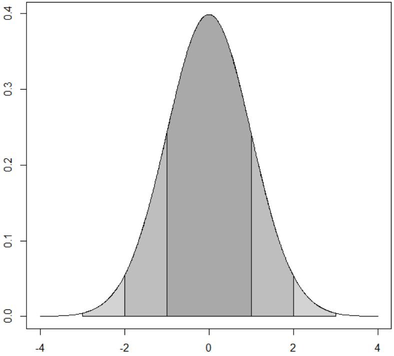


- Note that $\Pr(X \geq 2.2) = \Pr(X \leq 1.8)$!
- Thanks to symmetry, we have $\Pr(X \leq \mu - d) = \Pr(X \geq \mu + d)$ for all d if $X \sim \text{ND}(\mu, \sigma)$.

Quality control

- With probability 5%, a bag **does not pass the quality standard**, i.e., either too heavy or too light.
- Whether 5% is large depends.
- As long as the distribution is normal:

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



Cumulative distribution functions

- It is so often that we need to calculate the probability for a random variable to be smaller than a given value.
- For a random variable, we define

$$F(x) = \Pr(X \leq x)$$

as the **cumulative distribution functions** (cdf).

- In the previous example, we have $F(1.8) \approx 5\%$ and $F(2.2) \approx 95\%$.