Stat 154: Elementary Statistics

Jongyun Jung

Minnesota State University, Mankato

Ch 5: **Continuous Probability Distributions**

jongyun.jung@mnsu.edu

February 19, 2019

Overview

Definitions

- **Continuous Probability Distribution**: A continuous probability distribution satistifes three conditions:
  1. The random variable X is continuous.
  2. The probability density function (pdf) $f(x) \geq 0$ for all x.
  3. The total area under the pdf curve is **one**.
- **Uniform Distribution**: The general formula for the probability density function (pdf) of the uniform distribution is

$$f(x) = \frac{1}{B-A} \quad \text{for} \quad A \leq x \leq B \tag{1}$$

  - The mean of X is $\mu = \dfrac{A+B}{2}$ and the variance of X is $\sigma^2 = \dfrac{(B-A)^2}{12}$.
  - The probability density function for the standard uniform distribution is $f(x) = 1$ for $0 \leq x \leq 1$.

The Normal Distribution

- The general formula for the pdf of the normal distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \text{for} \quad -\infty < x < \infty \tag{2}$$

- where μ is the location parameter and σ is teh scale parameter.
- The mean of X is μ and the standard deviation of X is σ.
- When $\mu = 0$ and $\sigma = 1$, then it is called the **standard normal distribution**. The probability density function for the standard normal distribution is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \quad \text{for} \quad -\infty < z < \infty \tag{3}$$

- where $Z = \dfrac{X-\mu}{\sigma}$

The Normal Distribution

- By using the relation $Z = \dfrac{X - \mu}{\sigma}$, any normal random variable X can be converted to the standard normal random variable Z.
- Then **Table 2: Standard Normal Cumulative Probability** $P(Z \leq z)$ in **Appendix II** can be used to obtain the probabilities for any range of measurements.

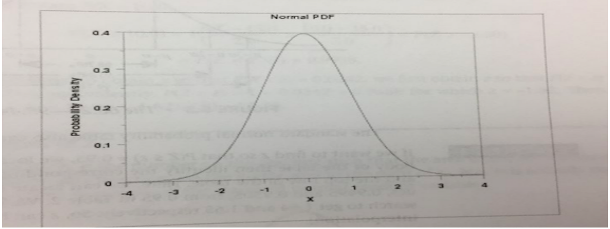


Figure: Density function of the standard normal

The Normal Distribution

EXAMPLE 5.1 Since $f(z)$ is symmetric about 0, $P(Z \le 1.5) = P(Z \ge -1.5) = 0.9332$.

$$P(Z \le -1.5) = 1 - P(Z \ge -1.5) = 1 - P(Z \le 1.5) = 1 - 0.9332 = 0.0668.$$

Since the measurements are continuous, $P(Z \le 1.5) = P(Z < 1.5) = 0.9332$, that is, $P(Z = 1.5) = 0$.

EXAMPLE 5.2 In a similar manner to the **Empirical Rule**, any normal random variable has the following three properties:

1. 68.26% of all possible values lies within one standard deviation about the mean.
2. 95.44% of all possible values lies within two standard deviations about the mean.
3. 99.74% of all possible values lies within three standard deviations about

Figure: The Normal Distribution

The Normal Distribution

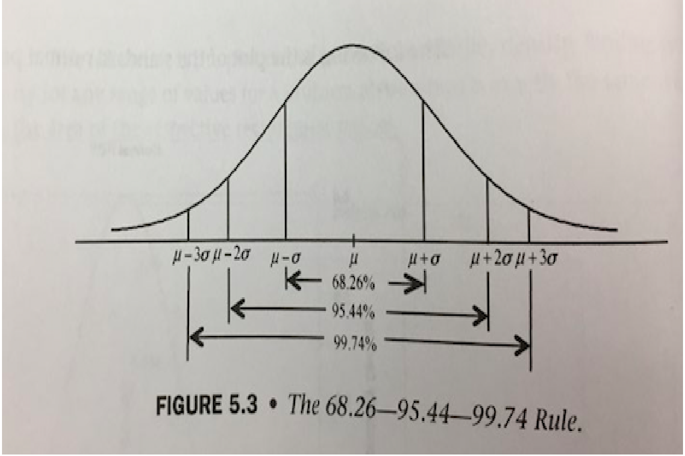


Figure: The 68.26 – 95.44 – 99.74 Rule

The Normal Distribution

EXAMPLE 5.5 Let us consider that X has a normal distribution with mean $\mu = 15.0$ and standard deviation $\sigma = 2.0$. In short, we can indicate it as $X \sim N(15.0, 2.0^2)$.

To obtain $P(X > 10.0)$,

$$P(X > 10.0) = P\left(\frac{X - 15.0}{2.0} > \frac{10.0 - 15.0}{2.0}\right) = P(Z > -2.50)$$

$$= P(Z < 2.50) = 0.9938.$$

Similarly, to obtain x so that $P(X < x) = 0.0342$, we first obtain z so that $P(Z < z) = 0.0342$; equivalently, $P(Z > z) = 1 - 0.0342 = 0.9658$ for which $z = -1.82$. Then $x = \mu + z\sigma = 15.0 + (-1.82)(2.0) = 11.36$.

Figure: The Normal Distribution

The Normal Distribution

Table 2: Standard Normal Cumulative Probability $P(Z \leq z)$ Table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5953
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

FIGURE 5.4 • *Finding z-values by the Standard Normal Table.*

Figure: Finding z-values by the Standard Normal Table

Notation of Z_α

- Z_α is the number on the real line such that the area under the standard normal density funciton, on the x-axis, and to the right of Z_α.

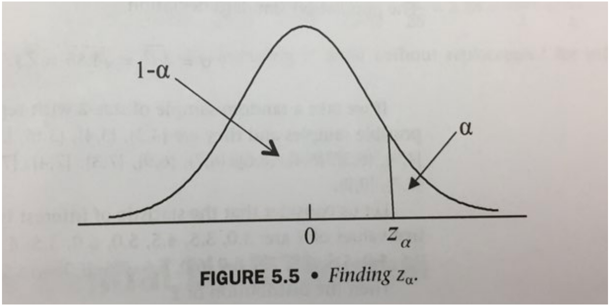
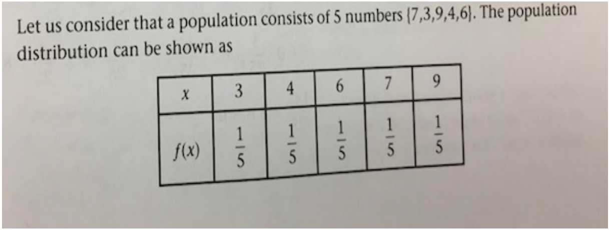


Figure: Finding Z_α

Sampling Distributions

- Let us consider T a statistic (a function of a sample). The distribution of T is known as the sampling distribution of T.



Let us consider that a population consists of 5 numbers {7,3,9,4,6}. The population distribution can be shown as

x	3	4	6	7	9
$f(x)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

Figure: Ex 5-7

- In general,

$$\mu_{\overline{x}} = \mu \text{ and } \sigma^2_{\overline{x}} = \frac{\sigma^2}{n}.$$

The Central Limit Theorem

- **Theorem 1**: Let us consider $X_1, X_2, ..., X_n$ as a random sample from a normal distribution with mean μ and variance σ^2. Then the sample mean $\overline{X}$ has a normal distribution with mean μ and variance $\dfrac{\sigma^2}{n}$.
  Here, a mean of $\overline{X}$, $E(\overline{X}) = \mu$ and a variance of $\overline{X}$, $V(\overline{X}) = \dfrac{\sigma^2}{n}$.
- **Theorem 2**: Let us consider $X_1, X_2, ..., X_n$ as a random sample from a normal distribution with mean μ and variance σ^2. Then the sample total $\sum X$ has a normal distribution with mean $n\mu$ and variance $n\sigma^2$. Here, a mean of $\sum X$, $E(\sum X) = n\mu$ and a variance of $\sum X$, $V(\sum X) = n\sigma^2$.

The Central Limit Theorem

- **Theorem 3**: Let us consider $X_1, X_2, ..., X_n$ as a random sample from a normal distribution with mean μ and variance σ^2. Then for a large $n(n \geq 30)$, the sample mean $\overline{X}$ has an approximately normal distribution with mean μ and variance $\frac{\sigma^2}{n}$. Here, a mean of $\overline{X}, E(\overline{X}) = \mu$ and a variance of $\overline{X}, V(\overline{X}) = \frac{\sigma^2}{n}$.
- **Theorem 4**: Let us consider $X_1, X_2, ..., X_n$ as a random sample from a normal distribution with mean μ and variance σ^2. Then for a large $n(n \geq 30)$, the sample total $\sum X$ has an approximately normal distribution with mean $n\mu$ and variance $n\sigma^2$. Here, a mean of $\sum X, E(\sum X) = n\mu$ and a variance of $\sum X, V(\sum X) = n\sigma^2$.

The Central Limit Theorem

EXAMPLE 5.9

Let X be a random variable that represents the level of glucose in the blood (milligrams per deciliter of blood) after a 12-hour fast. Assume that for people under 50 years old, X has a distribution that is approximately normal with mean 85 and standard deviation 25. A test result of $X < 40$ is an indication of severe excess insulin, and medication is usually prescribed.

a. What is the probability that, on a single test, $X < 40$?

b. Suppose a doctor uses the average $\bar{X}$ for two tests taken about a week apart. What can we say about the probability distribution of $\bar{X}$? What is the probability that $\bar{X} < 40$?

Solution:

a. $$P(X < 40) = P\left(\frac{X - 85}{25} < \frac{40 - 85}{25}\right) = P(Z < -1.80)$$

$$= 1 - 0.9641 = 0.0359.$$

b. By Theorem 1, the probability distribution of $\bar{X}$ is normal with mean 85 and standard deviation

$$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 25/\sqrt{2} = 17.68.$$

And

$$P(\bar{X} < 40) = P\left(\frac{\bar{X} - 85}{17.68} < \frac{40 - 85}{17.68}\right) = P(Z < -2.55)$$

$$= 1 - 0.9946 = 0.0054.$$

Figure: Ex 5-9

The Central Limit Theorem

EXAMPLE 5.10 A bottle of 5,000 ml wine was served at an end of semester party. There was a total of 48 guests attending the party. Assume that the amount of wine consumed by each guest followed some specific distribution with mean 100 ml and standard deviation 32 ml. We also assume that the guests drank the wine independently. What is the probability that the bottle was still not empty after all 48 guests had the wine?

Solution: Let X_i be the amount of wine that ith guest had consumed. By Theorem 4, the probability distribution of the total amount wine consumed by 48 guests, $\sum_{i=1}^{48} X_i$, is normal with mean 4,800 ml and standard deviation $\sqrt{48 \cdot 32^2} = 128\sqrt{3}$ ≈ 221.70 ml. So the probability that the bottle was not empty is actually

$$P\left(\sum_{i=1}^{48} X_i < 5000\right) = P\left(\frac{\sum_{i=1}^{48} X_i - 4800}{221.70} < \frac{5000 - 4800}{221.70}\right) \cong P(Z < 0.9021)$$

$$= 0.8165.$$

So, there was a very high chance that the wine was enough for the party.

Figure: Ex 5-10

Normal Approximation to the Binomial Distribution

- The **Central Limit Theorem** concept can also applied **to approximate discrete probability events by using continuous distribution such as the normal distribution.**
- The **binomial probabilities** can be **approximated** using the **normal distribution**.

Normal Approximation to the Binomial Distribution

LE 5.11 Let us consider that X has a binomial distribution with $n = 100$ and $p = 0.87$. We want to compute the probability that X is at least 90, that is, $P(X \geq 90)$.

The exact probability is

$$
\begin{aligned}
P(X \geq 90) &= P(X = 90) + P(X = 91) + \cdots + P(X = 100) \\
&= {}_{100}C_{90}(0.87)^{90}(0.13)^{10} + {}_{100}C_{91}(0.87)^{91}(0.13)^{9} + \cdots \\
&\quad + {}_{100}C_{100}(0.87)^{100}(0.13)^{0} \\
&= 0.0860 + 0.0632 + \cdots + 0.0000 = 0.2337.
\end{aligned}
$$

The same probability can be approximated using the normal distribution as $P(X \geq 90) \cong P(X \geq 89.5)$ after considering the **continuity correction**. (The variable as defined is discrete, but when we are using a normal table then we are using continuous scale of measurements; hence, a correction is needed, which is known as the **continuity correction**).

Here, X follows an approximate normal distribution with mean $\mu = np$ and standard deviation

$$\sigma = \sqrt{np(1 - p)}.$$

Then

$$
\begin{aligned}
P(X > 89.5) &= P\left(\frac{X - 100(0.87)}{\sqrt{100(0.87)(0.13)}} > \frac{89.5 - 100(0.87)}{\sqrt{100(0.87)(0.13)}}\right) \cong P(Z > 0.74) \\
&= 1 - P(Z \leq 0.74) = 1 - 0.7704 = 0.2296
\end{aligned}
$$

is close to the exact value 0.2337.

The normal approximation to the binomial probabilities is suggested when $np \geq 5$ and $n(1 - p) \geq 5$.

Figure: Normal Approximation to the Binomial Distribution

References

Mezbahur Rahman, Deepak Sanjel, Han Wu. Statistics Introduction, Revised Printing
KendallHunt