Math 408 - Mathematical Statistics

Lecture 17. The Normal Approximation to the Distribution of $\overline{X}_n$

March 1, 2013

Agenda

- Normal Approximation (theoretical result)
- Approximation of the Error Probabilities (application 1)
- Confidence Intervals (application 2)
- Example: Hospitals
- Summary

We previous Lectures, we found the mean and the variance of the sample mean:

$$\mathbb{E}[\overline{X}_n] = \mu \qquad \mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)$$

Ideally, we would like to know the **entire distribution** of $\overline{X}_n$ (sampling distribution) since it would tell us everything about the random variable $\overline{X}_n$

Reminder:

If $X_1, \ldots, X_n$ are i.i.d. with the common mean μ and variance σ^2, then the sample mean $\overline{X}_n$ has the following properties:

1. $\mathbb{E}[\overline{X}_n] = \mu, \qquad \mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$
2. CLT:

$$\mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) \to \Phi(z), \qquad \text{as } n \to \infty$$

   where $\Phi(z)$ is the CDF of $\mathcal{N}(0,1)$

Q: Can we use these results to obtain the distribution of $\overline{X}_n$?
A: No. In simple random sampling, X_i are not independent.
Moreover, it makes no sense to have n tend to infinity while N is fixed.

Nevertheless, it can be shown that if n is large, but still small relative to N, then $\overline{X}_n$ is **approximately normally distributed**

$$\boxed{\overline{X}_n \dot{\sim} \mathcal{N}(\mu, \sigma^2_{\overline{X}_n})} \qquad \sigma_{\overline{X}_n} = \frac{\sigma}{\sqrt{n}}\sqrt{1 - \frac{n-1}{N-1}}$$

How can we use this results?

Suppose we want to find the probability that the error made in estimating μ by $\overline{X}_n$ is less than $\varepsilon > 0$. In symbols, we want to find

$$\mathbb{P}(|\overline{X}_n - \mu| \leq \varepsilon) = ?$$

Theorem

From $\overline{X}_n \dot{\sim} \mathcal{N}(\mu, \sigma^2_{\overline{X}_n})$ *it follows that*

$$\boxed{\mathbb{P}(|\overline{X}_n - \mu| \leq \varepsilon) \approx 2\Phi\left(\frac{\varepsilon}{\sigma_{\overline{X}_n}}\right) - 1}$$

Confidence Intervals

Let $\alpha \in [0, 1]$

Definition

A $100(1-\alpha)\%$ **confidence interval** for a population parameter θ is a random interval calculated from the sample, which contains θ with probability $1-\alpha$.

Interpretation:

If we were to take many random samples and construct a confidence interval from each sample, then about $100(1-\alpha)\%$ of these intervals would contain θ.

Our goal: to construct a confidence interval for μ

Let z_α be that number such that the area under the standard normal density function to the right of z_α is α. In symbols, z_α is such that

$$\Phi(z_\alpha) = 1 - \alpha$$

Useful property:

$$z_{1-\alpha} = -z_\alpha$$

Confidence interval for μ

Theorem

An (approximate) $100(1-\alpha)\%$ confidence interval for μ is

$$\left(\overline{X}_n - z_{\frac{\alpha}{2}}\sigma_{\overline{X}_n}, \overline{X}_n + z_{\frac{\alpha}{2}}\sigma_{\overline{X}_n}\right)$$

That is the probability that μ lies in that interval is approximately $1-\alpha$

$$\boxed{\mathbb{P}(\overline{X}_n - z_{\frac{\alpha}{2}}\sigma_{\overline{X}_n} \leq \mu \leq \overline{X}_n + z_{\frac{\alpha}{2}}\sigma_{\overline{X}_n}) \approx 1-\alpha}$$

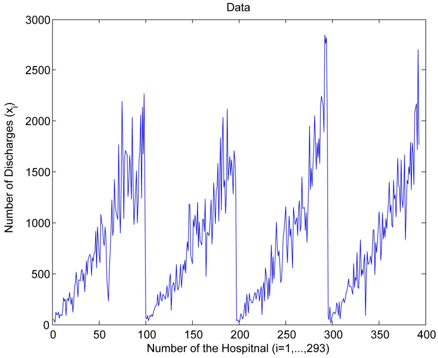
Remarks:

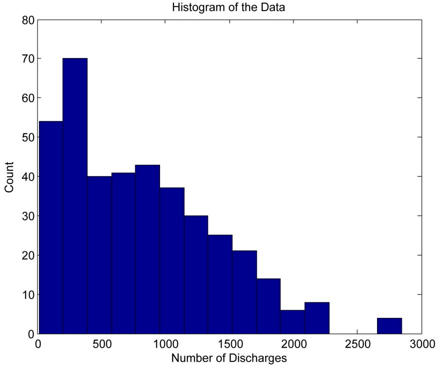
- This confidence interval is random. The probability that it covers μ is $(1-\alpha)$
- In practice, $\alpha = 0.1, 0.05, 0.01$ (depends on a particular application)
- Since $\sigma_{\overline{X}_n}$ is not known (it depends on σ), $s_{\overline{X}_n}$ is used instead of $\sigma_{\overline{X}_n}$

Example: Hospitals

Data: Herkson (1976):

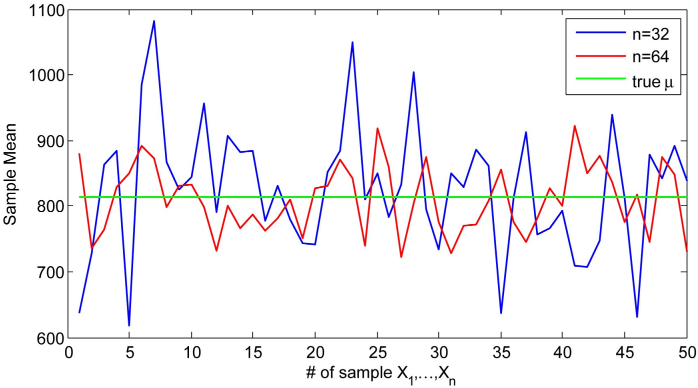
- The population consists of $N = 393$ short-stay hospitals
- Let x_i be the number of patients discharged from the i^{th} hospital during January 1968.





Example: Hospitals

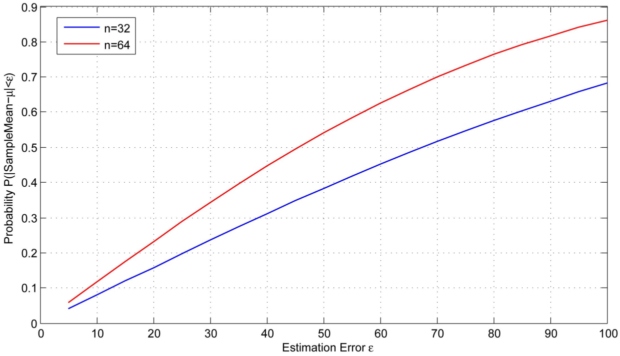
- Population mean $\mu = 814.6$, and population variance $\sigma^2 = (589.7)^2$
- Let us consider two case $n_1 = 32$ and $n_2 = 64$.

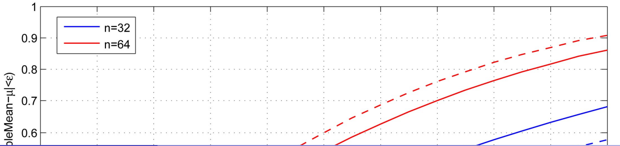


- True std of $\overline{X}_n$: $\sigma_{\overline{X}_n} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)}$, $\sigma_{\overline{X}_{32}} = 100$, $\sigma_{\overline{X}_{64}} = 67.5$

Example: Hospitals

$$\mathbb{P}(|\overline{X}_n - \mu| \le \varepsilon) \approx 2\Phi\left(\frac{\varepsilon}{\sigma_{\overline{X}_n}}\right) - 1$$



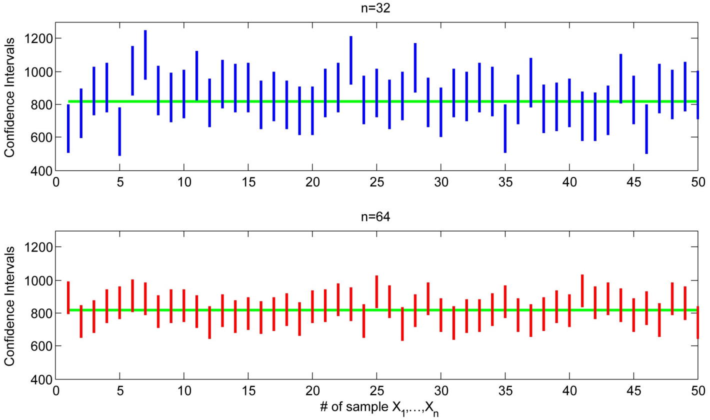


Example: Hospitals

$100(1-\alpha)\%$ confidence interval for μ is

$$\left(\overline{X}_n - z_{\frac{\alpha}{2}}\sigma_{\overline{X}_n}, \overline{X}_n + z_{\frac{\alpha}{2}}\sigma_{\overline{X}_n}\right)$$

$\alpha = 0.1$:



Interval width: 329.1 for $n = 32$ and 222.2 for $n = 64$

Summary

- The sample mean is approximately normal

$$\boxed{\overline{X}_n \dot{\sim} \mathcal{N}(\mu, \sigma^2_{\overline{X}_n})} \qquad \sigma_{\overline{X}_n} = \frac{\sigma}{\sqrt{n}}\sqrt{1 - \frac{n-1}{N-1}}$$

- Probability of error

$$\mathbb{P}(|\overline{X}_n - \mu| \leq \varepsilon) \approx 2\Phi\left(\frac{\varepsilon}{\sigma_{\overline{X}_n}}\right) - 1$$

- $100(1-\alpha)\%$ confidence interval for μ is

$$(\overline{X}_n - z_{\frac{\alpha}{2}}\sigma_{\overline{X}_n}, \overline{X}_n + z_{\frac{\alpha}{2}}\sigma_{\overline{X}_n})$$