50.034 - Introduction to Probability and Statistics

Week 6 – Lecture 12

January–May Term, 2019

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Lecture 12 Assigned Readings: 6.3, 6.4

Outline of Lecture

- Central Limit Theorem
- Convergence in distribution
- Normal approximations of binomial and Poisson distributions
- Correction for continuity

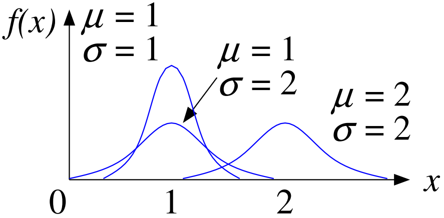
Recall: Normal distribution (Lecture 10)

A continuous R.V. X is called normal if its pdf is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

for some parameters μ and σ satisfying $-\infty < \mu < \infty$ and $\sigma > 0$.

- Its distribution is called the normal (or Gaussian) distribution.
- Many real-world R.V.'s have the normal distribution.
- $\mathbf{E}[X] = \mu$, and $\mathrm{var}(X) = \sigma^2$.
- The graph of its pdf is sometimes called a "bell-shaped curve".
  - This graph is symmetric about $x = \mu$, shaped like a "bell".



Sample mean of random sample of normal R.V.'s

Recall: A random sample is a collection $\{X_1, \ldots, X_n\}$ of **iid** R.V.'s.

Fact: (Lecture 11) If $\{X_1, \ldots, X_n\}$ is a random sample, such that each X_i has mean μ and variance σ^2, then the sample mean $\overline{X}_n$ has mean μ and variance $\frac{\sigma^2}{n}$.

Fact: (Lecture 10) If $X_1, \ldots, X_n$ are **independent** R.V.'s, such that each X_i has the **normal distribution** with mean μ_i and variance σ_i^2, then $a_1X_1 + \cdots + a_nX_n + b$ has the **normal distribution** with mean $a_1\mu_1 + \cdots + a_n\mu_n + b$ and variance $a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2$.

- Here, $a_1, \ldots, a_n, b$ are constants such that at least one of $a_1, \ldots, a_n$ is non-zero.

Consequence: If $\{X_1, \ldots, X_n\}$ is a random sample, such that each X_i has the **normal distribution** with mean μ and variance σ^2, then the sample mean satisfies $\overline{X}_n \sim N(\mu, \frac{\sigma^2}{n})$.

Question: What about random samples with other distributions?

Central Limit Theorem

Let $\{X_1, \ldots, X_n\}$ be a random sample, such that each X_i has mean μ and variance σ^2. Let $\overline{X}_n$ be its sample mean.

What we know so far:

- $\mathbf{E}[\overline{X}_n] = \mu$, $\mathrm{var}(\overline{X}_n) = \frac{\sigma^2}{n}$ (for any distribution).
- If each X_i is normal, then $\overline{X}_n$ is normal.

Theorem: (**Central Limit Theorem**) For every real number z,

$$\lim_{n\to\infty} \Pr\left(\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z\right) = \Phi(z),$$

where $\Phi(z)$ denotes the standard normal cdf.

- **Recall:** If $Y \sim N(\mu_Y, \sigma_Y^2)$, then $\frac{Y-\mu_Y}{\sigma_Y} \sim N(0,1)$.
- Hence, $\Pr(\frac{Y-\mu_Y}{\sigma_Y} \leq z) = \Phi(z)$.
- Abbreviation: CLT for Central Limit Theorem.

Interpretation of Central Limit Theorem

Let $\{X_1, \ldots, X_n\}$ be a random sample, such that each X_i has mean μ and variance σ^2. Let $\overline{X}_n$ be the sample mean.

- As n becomes larger, $\overline{X}_n$ becomes closer to a normal R.V. with mean μ and variance $\frac{\sigma^2}{n}$.
- In other words, as long as its sample size n is sufficiently large, then the normal distribution with the same mean μ and same variance $\frac{\sigma^2}{n}$ is a good approximation to the distribution of the sample mean $\overline{X}_n$.

Important Remarks:

- X_i can have **any** distribution, maybe "very different" from the normal distribution. It could be discrete, continuous, or mixed.
- Whatever distribution X_i has, the CLT says that the sample mean $\overline{X}_n$ (a R.V. with another distribution) has a distribution that is approximately normal.
- The CLT **does not** tell us anything new about the distribution of X_i.

Example 1

Example of a bimodal graph

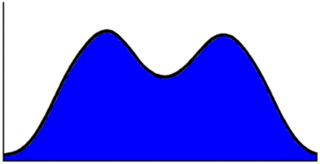


Image Source: Kathy Lam.

In several grasshopper species, female grasshoppers tend to be significantly larger than male grasshoppers.

Hence, if you measure the length of 1000 randomly selected grasshoppers, and plot a histogram of the number of observations (y-axis) versus length of grasshopper (x-axis), then the graph would be **bimodal**, i.e. has two peaks.

We have a random sample $\{X_1, \ldots, X_{1000}\}$, where each X_i (representing grasshopper length) has a bimodal distribution.

The CLT tells us that $\overline{X}_{1000}$ is approximately normal.

- In other words, the average grasshopper length, as you take different random samples of large sample sizes, would have a normal distribution, which in particular is **unimodal**.

Example 2

Coin Toss Experiment:
Toss a fair coin 10000 times. Let H be the total number of heads in these 10000 tosses. What is the probability $\Pr(H > 5120)$?

Solution: Let $\{X_1, \ldots, X_{10000}\}$ be a random sample, where $X_i = 1$ if the i-th toss yields heads, and $X_i = 0$ if the i-th toss yields tails.

- Every X_i is a Bernoulli R.V. with parameter 0.5.
- $\mathbb{E}[X_i] = 0.5$, $\mathrm{var}(X_i) = 0.25$.
- $\mathbf{E}[\overline{X}_{10000}] = 0.5$, $\mathrm{var}(\overline{X}_{10000}) = \frac{0.25}{10000} = 0.000025$.
- Standard deviation of $\overline{X}_{10000} = \sqrt{0.000025} = 0.005$.

By the central limit theorem, $\overline{X}_{10000}$ approximately has a normal distribution with mean 0.5 and variance 0.000025.

Example 2 (continued)

Since $H = X_1 + \cdots + X_{10000} = 10000 \cdot \overline{X}_{10000}$, it follows that

$$\Pr(H > 5120) = 1 - \Pr(H \leq 5120) = 1 - \Pr(\overline{X}_{10000} \leq \tfrac{5120}{10000} = 0.512)$$

$$= 1 - \Pr\left(\frac{\overline{X}_{10000} - 0.5}{0.005} \leq \frac{0.512 - 0.5}{0.005}\right)$$

$$\approx 1 - \Phi(\tfrac{0.512-0.5}{0.005}) = 1 - \Phi(2.4),$$

where $\Phi(z)$ denotes the standard normal cdf.
From the table, $\Phi(2.40) \approx 0.9918$. Therefore,

$$\Pr(H > 5120) \approx 1 - \Phi(2.4) \approx 1 - 0.9918 = 0.0082.$$

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{(2\pi)^{1/2}} \exp\left(-\tfrac{1}{2}u^2\right) du$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	0.60	0.7257	1.20	0.8849	1.80	0.9641	2.40	0.9918
0.01	0.5040	0.61	0.7291	1.21	0.8869	1.81	0.9649	2.41	0.9920
0.02	0.5080	0.62	0.7324	1.22	0.8888	1.82	0.9656	2.42	0.9922
0.03	0.5120	0.63	0.7357	1.23	0.8907	1.83	0.9664	2.43	0.9925
0.04	0.5160	0.64	0.7389	1.24	0.8925	1.84	0.9671	2.44	0.9927
0.05	0.5199	0.65	0.7422	1.25	0.8944	1.85	0.9678	2.45	0.9929
0.06	0.5239	0.66	0.7454	1.26	0.8962	1.86	0.9686	2.46	0.9931
0.07	0.5279	0.67	0.7486	1.27	0.8980	1.87	0.9693	2.47	0.9932

CLT in terms of sums of iid R.V.'s

Theorem: Let $X_1, X_2, X_3, \dots$ be an infinite sequence of iid R.V.'s, such that each X_i has mean μ and variance σ^2. For each integer n, define the R.V. $Y_n = X_1 + \cdots + X_n$. Then for every real number z,

$$\lim_{n\to\infty} \Pr\left(\frac{Y_n - n\mu}{\sqrt{n}\sigma} \le z\right) = \Phi(z),$$

where $\Phi(z)$ denotes the standard normal cdf.

Proof: The sample mean $\overline{X}_n$ satisfies $\overline{X}_n = \frac{Y_n}{n}$, hence

$$\Pr\left(\frac{Y_n - n\mu}{\sqrt{n}\sigma} \le z\right) = \Pr\left(\frac{n\overline{X}_n - n\mu}{\sqrt{n}\sigma} \le z\right) = \Pr\left(\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \le z\right).$$

So by the central limit theorem,

$$\lim_{n\to\infty} \Pr\left(\frac{Y_n - n\mu}{\sqrt{n}\sigma} \le z\right) = \lim_{n\to\infty} \Pr\left(\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \le z\right) = \Phi(z).$$

Remark: Conversely, we could instead start with this theorem and prove the CLT. Thus, the given theorem is actually a reformulation of the CLT.

Reinterpretation of Central Limit Theorem

Let $X_1, X_2, X_3, \dots$ be an infinite sequence of iid R.V.'s, such that each X_i has mean μ and variance σ^2. For each integer n, let Y_n be the sum $Y_n = X_1 + \cdots + X_n$.

- As n becomes larger, Y_n becomes closer to a normal R.V. with mean $n\mu$ and variance $n\sigma^2$.
- In other words, as long as n is sufficiently large, then the normal distribution with the same mean $n\mu$ and same variance $n\sigma^2$ is a good approximation to the distribution of Y_n.

Important Remarks:

- X_i can have **any** distribution, maybe "very different" from the normal distribution. It could be discrete, continuous, or mixed.
- Whatever distribution X_i has, the CLT says that the sum Y_n (a R.V. with another distribution) has a distribution that is approximately normal.
- The CLT **does not** tell us anything new about the distribution of X_i.

Revisited Example: Average age in Singapore

(Slight variant of Lecture 11 Example 3)

Let μ be the average age (in years) of all 5.6 million people in Singapore. We do not know what the exact value of μ is, and suppose we want to get an estimate of the value of μ.

We decide to conduct a poll on n randomly selected people to record their ages. Suppose it is known that the age of any randomly selected person in Singapore, treated as a R.V., has a standard deviation of 20 years.

What should the value of n be so that the average age of these n polled people is within 2 years of the actual average age (μ years), with probability **approximately** at least 0.99?

Revisited Example: Average age in Singapore (continued)

Let X_i be the age (in years) of the i-th person polled.
We are given that $\{X_1, \dots, X_n\}$ forms a random sample.
We are also given that each X_i has standard deviation $\sigma = 20$.

We want to determine the minimum possible value for n, so that the sample mean $Y = \frac{X_1 + \cdots + X_n}{n}$ satisfies

$$\Pr(|Y - \mu| < 2) \geq 0.99.$$

In Lecture 11 (Example 3), we applied **Chebyshev's inequality** and concluded that polling $n = 10000$ people would **guarantee** that $\Pr(|Y - \mu| < 2) \geq 0.99$.

We could also apply the **central limit theorem** and find the smallest value for n that would satisfy $\Pr(|Y - \mu| < 2) \geq 0.99$ **approximately**.

Revisited Example: Average age in Singapore (continued)

By the central limit theorem, Y approximately has the normal distribution with mean μ and variance $\frac{20^2}{n}$, hence

$$Z = \frac{Y - \mu}{\frac{20}{\sqrt{n}}} = \frac{(Y - \mu)\sqrt{n}}{20}$$

satisfies $Z \sim N(0, 1)$ approximately.
Note that

$$\begin{aligned}
\Pr(|Y - \mu| < 2) &= \Pr\left(\frac{-2\sqrt{n}}{20} < \frac{(Y - \mu)\sqrt{n}}{20} < \frac{2\sqrt{n}}{20}\right) \\
&= \Pr(-0.1\sqrt{n} < Z < 0.1\sqrt{n}) \\
&\approx \Phi(0.1\sqrt{n}) - \Phi(-0.1\sqrt{n}) \\
&= \Phi(0.1\sqrt{n}) - (1 - \Phi(0.1\sqrt{n})) \\
&= 2 \cdot \Phi(0.1\sqrt{n}) - 1,
\end{aligned}$$

where $\Phi(z)$ denotes the standard normal cdf.

Revisited Example: Average age in Singapore (continued)

We want to find a value of n such that $\Pr(|Y-\mu|<2)\geq 0.99$ approximately, i.e. it suffices to find a value of n such that

$$2\cdot\Phi(0.1\sqrt{n})-1\geq 0.99,$$

or equivalently, $\Phi(0.1\sqrt{n})\geq 0.995$.
From the table, the closest value we can find for z satisfying $\Phi(z)=0.995$ is $z=2.58$.
Therefore $0.1\sqrt{n}\geq 2.58$, which implies $n\geq\frac{1}{(0.1)^2}\times 2.58^2\approx 666$.

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	0.60	0.7257	1.20	0.8849	1.80	0.9641	2.40	0.9918
0.01	0.5040	0.61	0.7291	1.21	0.8869	1.81	0.9649	2.41	0.9920
0.02	0.5080	0.62	0.7324	1.22	0.8888	1.82	0.9656	2.42	0.9922
0.03	0.5120	0.63	0.7357	1.23	0.8907	1.83	0.9664	2.43	0.9925
0.04	0.5160	0.64	0.7389	1.24	0.8925	1.84	0.9671	2.44	0.9927
0.05	0.5199	0.65	0.7422	1.25	0.8944	1.85	0.9678	2.45	0.9929
0.06	0.5239	0.66	0.7454	1.26	0.8962	1.86	0.9686	2.46	0.9931
0.07	0.5279	0.67	0.7486	1.27	0.8980	1.87	0.9693	2.47	0.9932
0.08	0.5319	0.68	0.7517	1.28	0.8997	1.88	0.9699	2.48	0.9934
0.09	0.5359	0.69	0.7549	1.29	0.9015	1.89	0.9706	2.49	0.9936
0.10	0.5398	0.70	0.7580	1.30	0.9032	1.90	0.9713	2.50	0.9938
0.11	0.5438	0.71	0.7611	1.31	0.9049	1.91	0.9719	2.52	0.9941
0.12	0.5478	0.72	0.7642	1.32	0.9066	1.92	0.9726	2.54	0.9945
0.13	0.5517	0.73	0.7673	1.33	0.9082	1.93	0.9732	2.56	0.9948
0.14	0.5557	0.74	0.7704	1.34	0.9099	1.94	0.9738	2.58	0.9951
0.15	0.5596	0.75	0.7734	1.35	0.9115	1.95	0.9744	2.60	0.9953

Chebyshev's Inequality vs Central Limit Theorem

In the previous example (on average age in Singapore):

We used Chebyshev's inequality to conclude that polling $n = 10000$ people would **guarantee** that $\Pr(|Y - \mu| < 2) \geq 0.99$ is satisfied.

- Information used: Every X_i has standard deviation 20.

We used the CLT to conclude that polling $n = 666$ people would imply $\Pr(|Y - \mu| < 2) \geq 0.99$ is satisfied **approximately**.

- Information used: Every X_i has standard deviation 20, and $Y = \overline{X}_n$ is the sample mean of $\{X_1, \ldots, X_n\}$.

Remarks:

- Chebyshev's inequality is a crude inequality that can be applied to **any** R.V.
- If we want an approximation (instead of a guaranteed bound), and if the R.V. represents a sample mean (or a sum of iid R.V.'s), then we can use CLT to get a better approximation.

Convergence in distribution

Let $X_1, X_2, X_3, \dots$ be an infinite sequence of R.V.'s, and for each i, let $F_i(x)$ denote the cdf of X_i.

Definition: Suppose X is a R.V. with cdf $F(x)$. We say that the sequence $X_1, X_2, X_3, \dots$ converges in distribution to X if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for all $x \in \mathbb{R}$ at which F is continuous.

- More simply, we say X_n converges in distribution to X.
- We usually write $X_n \xrightarrow{\text{d}} X$.
- The distribution of X is called the asymptotic distribution of the sequence $X_1, X_2, X_3, \dots$

Same technicality as in other kinds of convergence previously seen:
When we are given a finite sequence $X_1, \dots, X_n$ of **iid** R.V.'s, it still makes sense to take the limit $n \to \infty$. Since every X_i has the exact same distribution, we could extend our sequence to an infinite sequence $X_1, X_2, \dots$ of independent R.V.'s, with each R.V. having this same distribution.

Remarks on convergence in distribution

Interpretation of convergence in distribution:
"X_n converges in distribution to X" means that if we consider the cdf $F_n(x)$ of X_n, as well as the cdf $F(x)$ of X, then for sufficiently large n, we have $F_n(x_0) \approx F(x_0)$ for every real number x_0 such that $F(x)$ is continuous at $x = x_0$.

Fact: $X_n \xrightarrow{\text{p}} X$ implies $X_n \xrightarrow{\text{d}} X$, but not conversely.

Equivalent terminology: The following mean exactly the same.

- X_n converges in distribution to X.
- X_n converges weakly to X.
- X_n converges in law to X.

Remark: The weak law of large numbers (Lecture 11) is stated in terms of convergence in probability, and NOT in terms of weak convergence (i.e. NOT in terms of convergence in distribution).

Convergence under continuous functions

Convergence in probability:

Theorem: (Lecture 11) Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of R.V.'s, let X be a R.V., and suppose that $X_n \xrightarrow{\text{p}} X$. If $h(t)$ is a real-valued continuous function on the reals, then $h(X_n) \xrightarrow{\text{p}} h(X)$.

Almost sure convergence:

Theorem: Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of R.V.'s, let X be a R.V., and suppose that $X_n \xrightarrow{\text{a.s.}} X$. If $h(t)$ is a real-valued continuous function on the reals, then $h(X_n) \xrightarrow{\text{a.s.}} h(X)$.

Convergence in distribution:

Theorem: Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of R.V.'s, let X be a R.V., and suppose that $X_n \xrightarrow{\text{d}} X$. If $h(t)$ is a real-valued continuous function on the reals, then $h(X_n) \xrightarrow{\text{d}} h(X)$.

These three theorems are collectively (i.e. together) called the **continuous mapping theorem**.

CLT in terms of convergence in distribution

Theorem: (Reformulation of Central Limit Theorem)
Suppose $\{X_1, \dots, X_n\}$ is a random sample, such that each X_i has mean μ and variance σ^2. Let $\overline{X}_n$ be its sample mean, and define $Z_n = \frac{(\overline{X}_n - \mu)\sqrt{n}}{\sigma}$. Then the sequence $Z_1, Z_2, Z_3, \dots$ converges in distribution to $Z \sim N(0,1)$, i.e. $Z_n \xrightarrow{\text{d}} Z$.

- In other words, the **asymptotic distribution** of the sequence $Z_1, Z_2, Z_3, \dots$ is the standard normal distribution.

Theorem: (Another reformulation of Central Limit Theorem)
Let $X_1, X_2, X_3, \dots$ be an infinite sequence of iid R.V.'s, such that each X_i has mean μ and variance σ^2. For each integer n, let Y_n be the sum $Y_n = X_1 + \cdots + X_n$, and define $Z_n = \frac{Y_n - n\mu}{\sqrt{n}\sigma}$. Then the sequence $Z_1, Z_2, Z_3, \dots$ converges in distribution to $Z \sim N(0,1)$, i.e. $Z_n \xrightarrow{\text{d}} Z$.

Other versions of the Central Limit Theorem

Note: There are several kinds of "central limit theorems"!

- These are generalizations of the original CLT, stated in terms of convergence in distribution, with certain conditions relaxed.

Theorem: (**Lyapunov's Central Limit Theorem**)
Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of independent R.V.'s, such that each X_i has mean μ_i and variance σ_i^2. For each integer n, define the R.V.

$$Z_n = \frac{\sum_{i=1}^{n}(X_i - \mu_i)}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^{0.5}}.$$

If every X_i satisfies $\mathbf{E}[|X_i - \mu_i|^3] < \infty$, and if

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} \mathbf{E}[|X_i - \mu_i|^3]}{(\sum_{i=1}^{n} \sigma_i^2)^{1.5}} = 0,$$

then the sequence $Z_1, Z_2, Z_3, \ldots$ converges in distribution to $Z \sim N(0,1)$, i.e. $Z_n \xrightarrow{d} Z$.

Approximations of binomial distributions

Suppose $0 < p < 1$, and let $X_1, X_2, X_3, \ldots$ be an infinite sequence of **independent Bernoulli R.V.'s**, each with parameter p.
For each integer n, let Y_n be the sum $Y_n = X_1 + \cdots + X_n$.

Recall: (Lecture 6) Y_n is a binomial R.V. with parameters n and p.

- Note: $\mathbf{E}[Y_n] = np$, $\mathrm{var}(Y_n) = np(1-p)$.

Consequence of CLT: If n is large, then Y_n is approximately a normal R.V. with the same mean np and same variance $np(1-p)$.

- In other words, the normal distribution is a good approximation to the binomial distribution if n is large.

Question: In Lecture 6, we saw that the Poisson distribution is a good approximation to the binomial distribution, at least under certain conditions. How are these two approximations different?

Poisson vs normal vs binomial distributions

Theorem: (Lecture 6) Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of R.V.'s such that each X_k is binomial with parameters n_k and p_k. Suppose $\lim_{k\to\infty} n_k = \infty$, $\lim_{k\to\infty} p_k = 0$, and $\lim_{k\to\infty} n_k p_k = \lambda$ for some value $\lambda > 0$. Then when k is sufficiently large, the cdf of X_k is approximately the cdf of the Poisson R.V. with parameter λ.

- In other words, $X_k \xrightarrow{\text{d}} X$, where X is a Poisson R.V. with parameter λ.

In contrast, the CLT gives the following:

Theorem: Given a **fixed p** $(0 < p < 1)$, let $X_1, X_2, X_3, \ldots$ be an infinite sequence of R.V.'s such that each X_k is binomial with parameters n_k and p (fixed). Suppose $\lim_{k\to\infty} n_k = \infty$. Then when k is sufficiently large, the cdf of X_k is approximately the cdf of the normal R.V. with mean np and variance $np(1-p)$.

- $n \to \infty$, $p \to 0$, np fixed $\Rightarrow$ binomial $\approx$ Poisson.
- $n \to \infty$, p fixed $\Rightarrow$ binomial $\approx$ normal.

Is the Poisson distribution approximately normal?

Question: So is Poisson distribution $\approx$ normal distribution?
Answer: Yes, if the parameter of the Poisson distribution is large.

Theorem: Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of R.V.'s such that each X_k is Poisson with parameter k. Then for every real number z,

$$\lim_{k\to\infty} \Pr\left(\frac{X_k - k}{\sqrt{k}} \leq z\right) = \Phi(z),$$

where $\Phi(z)$ denotes the standard normal cdf.

- In other words, $\frac{X_k - k}{\sqrt{k}} \xrightarrow{\text{d}} Z$, where Z is the standard normal R.V., i.e. $Z \sim N(0,1)$.

Interpretation: If λ is sufficiently large (e.g. $\lambda > 1000$), then the Poisson distribution with parameter λ is approximately the normal distribution with mean λ and variance λ.

Approximations of discrete R.V.'s by continuous R.V.'s

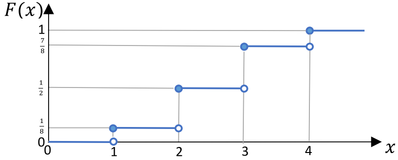
Discrete R.V.'s approximated by normal R.V.'s:

- A binomial R.V. X (with parameters n large, p fixed) can be approximated by $Z \sim N(np, np(1-p))$.
- A Poisson R.V. Y (with large mean λ) can be approximated by $Z' \sim N(\lambda, \lambda)$.

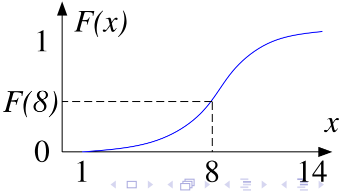
Problem: There are some inherent difficulties when discrete R.V.'s are approximated by continuous R.V.'s.

- e.g. $\Pr(X \leq 8) = \Pr(X < 9)$, but $\Pr(Z \leq 8) \neq \Pr(Z < 9)$.
- e.g. $\Pr(Y = 5) > 0$, but $\Pr(Z' = 5) = 0$.
- The cdf's (discrete vs continuous) have different properties.

cdf of discrete R.V.



cdf of continuous R.V.



Correction for Continuity

Goal: Introduce "corrections" so as to improve the approximations of discrete R.V.'s by continuous R.V.'s.

Main Idea: If X is a discrete R.V. taking only integer values, then for any integer k,

$$\Pr(X = k) = \Pr(k - 0.5 \leq X \leq k + 0.5).$$

So if Y is a continuous R.V. that approximates X, then we use

$$\Pr(X = k) \approx \Pr(k - 0.5 \leq Y \leq k + 0.5)$$

to compute an approximate value for $\Pr(X = k)$.

Similarly, $\Pr(X < k) = \Pr(X \leq k + 1)$, so we use

$$\Pr(X < k) = \Pr(X \leq k + 1) \approx \Pr(Y \leq k + 0.5)$$

to compute an approximate value for $\Pr(X < k) = \Pr(X \leq k + 1)$.

Correction for Continuity

Assumptions:

- X is a discrete R.V. taking only integer values
- Y is a continuous R.V. that approximates X.
- k is an integer.

Value	Approximation with continuity correction
$\Pr(X = k)$	$\Pr(k - 0.5 \leq Y \leq k + 0.5)$
$\Pr(X > k)$	$\Pr(Y \geq k + 0.5)$
$\Pr(X < k)$	$\Pr(Y \leq k - 0.5)$
$\Pr(X \geq k)$	$\Pr(Y \geq k - 0.5)$
$\Pr(X \leq k)$	$\Pr(Y \leq k + 0.5)$

Key advantage of continuity correction: Better approximations.

- If λ is sufficiently large (e.g. $\lambda > 1000$), then the normal distribution with mean λ and variance λ is a "good" approximation of the Poisson distribution with parameter λ.
- If we use the correction for continuity for approximating the cdf, then $\lambda > 10$ is considered "sufficiently large".

Example 3

Suppose that internet users visit the SUTD main website following a Poisson distribution with a rate of 30 visitors per hour.
What is the probability that there will be at most 700 visitors in a particular 24-hour period?

Solution: Let X be the number of visitors per 24-hour period. Since the number of visitors per hour follows a Poisson distribution with parameter 30, we infer that X is a Poisson R.V. with parameter $30 \times 24 = 720$. The pmf of X is:

$$p(x) = \frac{720^x e^{-720}}{x!}.$$

Let $Y \sim N(720, 720)$. Note that Y is an approximation of X.

Exact computation using pmf:

$$\Pr(X \leq 700) = \sum_{k=0}^{700} p(k) \approx 0.234595.$$

Example 3 (continued)

Exact computation using pmf:

$$\Pr(X \leq 700) = \sum_{k=0}^{700} p(k) \approx 0.234595.$$

We have defined that $Y \sim N(720, 720)$.
Let $Z = \frac{Y-720}{\sqrt{720}}$, and note that $Z \sim N(0,1)$.

Approximation without correction for continuity:

$$\Pr(X \leq 700) \approx \Pr(Y \leq 700) = \Pr\left(\frac{Y-720}{\sqrt{720}} \leq \frac{-20}{\sqrt{720}}\right)$$
$$\approx \Pr(Z \leq -0.745356) = \Phi(-0.745356) \approx 0.228028.$$

Approximation error: ≈ 0.006567 (i.e. roughly 0.66% error).

Approximation with correction for continuity:

$$\Pr(X \leq 700) \approx \Pr(Y \leq 700.5) = \Pr\left(\frac{Y-720}{\sqrt{720}} \leq \frac{-19.5}{\sqrt{720}}\right)$$
$$= \Pr(Z \leq -0.726722) = \Phi(-0.726722) \approx 0.233698.$$

Approximation error: ≈ 0.000897 (i.e. roughly 0.09% error).

Summary

- Central Limit Theorem
- Convergence in distribution
- Normal approximations of binomial and Poisson distributions
- Correction for continuity

Announcement:
The **mid-term exam** will be held in Week 8 (Wednesday, 2–4pm), at the **Multi-purpose hall**.

- Tested on all materials from Lectures 1–12 and Cohort classes weeks 1–6.
- This is Week 6. Today's lecture is Lecture 12.
- Lecture 14 (Week 8 Tuesday) will be a review lecture.