ESE 402/542 Recitation 3: Basic Tools of Estimation

Last Time...

- MGF and how to use it to derive moments of a random variable

Last Time...

- MGF and how to use it to derive moments of a random variable
- Some basics on CLT–visually and intro to usage

Today

- CLT examples and practice

CLT 1: Applying transforms before CLT

- **Scenario:** we have an object with initial mass m_0. It is then periodically subjected to impacts that leave a proportion of the mass left. Let's say there were a total of n impacts, each leaving $X_i \in (0.1, 1)$ proportion of the mass, where X_i are iid. We want to understand the distribution of the remaining mass after n impacts, Z_n.

CLT 1: Applying transforms before CLT

- **Scenario:** we have an object with initial mass m_0. It is then periodically subjected to impacts that leave a proportion of the mass left. Let's say there were a total of n impacts, each leaving $X_i \in (0.1, 1)$ proportion of the mass, where X_i are iid. We want to understand the distribution of the remaining mass after n impacts, Z_n.
- How can we transform the problem such that we can apply CLT? (Hint: We apply CLT to sums of things...how can we get sums from this?)

CLT 1: Applying transforms before CLT

- $Z_n = X_n \cdot X_{n-1} \ldots X_2 \cdot X_1 \cdot m_0$

CLT 1: Applying transforms before CLT

- $Z_n = X_n \cdot X_{n-1} \ldots X_2 \cdot X_1 \cdot m_0$
- Try out: logarithms

$$\begin{aligned}\log(Z_n) &= \log(X_n \cdot X_{n-1} \ldots X_2 \cdot X_1 \cdot m_0) \\ &= \log(m_0) + \sum_{i=1}^{n} \log(X_i)\end{aligned}$$

CLT 1: Applying transforms before CLT

- $Z_n = X_n \cdot X_{n-1} \ldots X_2 \cdot X_1 \cdot m_0$
- Try out: logarithms

$$\begin{aligned} \log(Z_n) &= \log(X_n \cdot X_{n-1} \ldots X_2 \cdot X_1 \cdot m_0) \\ &= \log(m_0) + \sum_{i=1}^{n} \log(X_i) \end{aligned}$$

- Check assumptions of CLT:

CLT 1: Applying transforms before CLT

- $Z_n = X_n \cdot X_{n-1} \ldots X_2 \cdot X_1 \cdot m_0$
- Try out: logarithms

$$\begin{aligned} \log(Z_n) &= \log(X_n \cdot X_{n-1} \ldots X_2 \cdot X_1 \cdot m_0) \\ &= \log(m_0) + \sum_{i=1}^{n} \log(X_i) \end{aligned}$$

- Check assumptions of CLT:
  1. $\log(X_i)$ are iid. Yep.

CLT 1: Applying transforms before CLT

- $Z_n = X_n \cdot X_{n-1} \ldots X_2 \cdot X_1 \cdot m_0$
- Try out: logarithms

$$\begin{aligned} \log(Z_n) &= \log(X_n \cdot X_{n-1} \ldots X_2 \cdot X_1 \cdot m_0) \\ &= \log(m_0) + \sum_{i=1}^{n} \log(X_i) \end{aligned}$$

- Check assumptions of CLT:
  1. $\log(X_i)$ are iid. Yep.
  2. $\operatorname{var}(\log(X_i)) < \infty$. Lower bound on $\log(X_i)$ is $\log(0.1) = -1$, upper bound is $\log(1) = 0$. Bounded rvs have finite variance (intuitive, but check Popoviciu's inequality if interested).
- Good to go!

CLT 1: Applying transforms before CLT

- As $n \to \infty$, $\log(Z_n)$ is distributed approximately normally.

CLT 1: Applying transforms before CLT

- As $n \to \infty$, $\log(Z_n)$ is distributed approximately normally.
- Side note: Z_n is said to have (approximately) log-normal distribution.

CLT 1: Applying transforms before CLT

- As $n \to \infty$, $\log(Z_n)$ is distributed approximately normally.
- Side note: Z_n is said to have (approximately) log-normal distribution.
- Problem source: Rice Chapter 5 Example G

CLT 2: Hypothesis Testing

- **Scenario:** Want to study whether "home-field advantage" exists at a statistically significant level for Major League Baseball games.

CLT 2: Hypothesis Testing

- **Scenario:** Want to study whether "home-field advantage" exists at a statistically significant level for Major League Baseball games.
- Data: 2431 games played, 1308 won on home turf.

CLT 2: Hypothesis Testing

- **Scenario:** Want to study whether "home-field advantage" exists at a statistically significant level for Major League Baseball games.
- Data: 2431 games played, 1308 won on home turf.
- What main assumptions do we need to even consider using CLT?

CLT 2: Hypothesis Testing

- **Scenario:** Want to study whether "home-field advantage" exists at a statistically significant level for Major League Baseball games.
- Data: 2431 games played, 1308 won on home turf.
- What main assumptions do we need to even consider using CLT?
  - "Many" samples (games). Sure.

CLT 2: Hypothesis Testing

- **Scenario:** Want to study whether "home-field advantage" exists at a statistically significant level for Major League Baseball games.
- Data: 2431 games played, 1308 won on home turf.
- What main assumptions do we need to even consider using CLT?
  - "Many" samples (games). Sure.
  - Games are iid...probably not exactly, but let's say it's close enough.

CLT 2: Hypothesis Testing

- **Scenario:** Want to study whether "home-field advantage" exists at a statistically significant level for Major League Baseball games.
- Data: 2431 games played, 1308 won on home turf.
- What main assumptions do we need to even consider using CLT?
  - "Many" samples (games). Sure.
  - Games are iid...probably not exactly, but let's say it's close enough.
  - Finite variance. Each game is either won (1) or lost (0) – Bernoulli, which has finite variance

CLT 2: Hypothesis Testing

- Form a null hypothesis: there is no home-field advantage. Alternate hypothesis: home-field advantage exists. Set significance level $\alpha = 0.05$

CLT 2: Hypothesis Testing

- Form a null hypothesis: there is no home-field advantage. Alternate hypothesis: home-field advantage exists. Set significance level $\alpha = 0.05$
- $H_0 : p = 0.5$, $H_1 : p > 0.5$

CLT 2: Hypothesis Testing

- Form a null hypothesis: there is no home-field advantage. Alternate hypothesis: home-field advantage exists. Set significance level $\alpha = 0.05$
- $H_0 : p = 0.5$, $H_1 : p > 0.5$
- Only want one-sided hypothesis. Reject null hypothesis if $p < \hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. $n = 2431$, 1308 won.

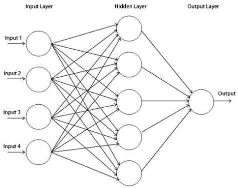
CLT 2: Hypothesis Testing

- Form a null hypothesis: there is no home-field advantage. Alternate hypothesis: home-field advantage exists. Set significance level $\alpha = 0.05$
- $H_0 : p = 0.5$, $H_1 : p > 0.5$
- Only want one-sided hypothesis. Reject null hypothesis if $p < \hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. $n = 2431$, 1308 won.
- Crunch the numbers! $\hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx 0.5215 > 0.5$. Reject the null hypothesis. There is statistical significant evidence that home-field advantage exists (under our assumptions).

CLT 2: Hypothesis Testing

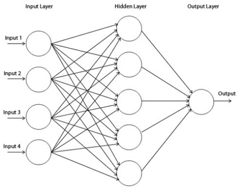
- Form a null hypothesis: there is no home-field advantage. Alternate hypothesis: home-field advantage exists. Set significance level $\alpha = 0.05$
- $H_0 : p = 0.5$, $H_1 : p > 0.5$
- Only want one-sided hypothesis. Reject null hypothesis if $p < \hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. $n = 2431$, 1308 won.
- Crunch the numbers! $\hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx 0.5215 > 0.5$. Reject the null hypothesis. There is statistical significant evidence that home-field advantage exists (under our assumptions).
- Source: Anderson, 2014

CLT 3: An Application to Deep Learning



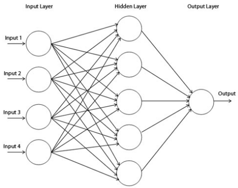
- Scenario: want to know behavior of randomly initialized one-layer neural net. Input $x = [x_1, \ldots, x_k]$ is dimension k, n hidden neurons, which each take value $\mathrm{ReLU}(\sum_{j=1}^{k} w_{ij} x_j)$. Output is number $f(x) = \sum_{i=1}^{n} a_i \mathrm{ReLU}(\sum_{j=1}^{k} w_{ij} x_j)$

CLT 3: An Application to Deep Learning



- Scenario: want to know behavior of randomly initialized one-layer neural net. Input $x = [x_1, \dots, x_k]$ is dimension k, n hidden neurons, which each take value $\mathrm{ReLU}(\sum_{j=1}^{k} w_{ij} x_j)$. Output is number $f(x) = \sum_{i=1}^{n} a_i \mathrm{ReLU}(\sum_{j=1}^{k} w_{ij} x_j)$
- Neural net is randomly initialized with $a_i \sim \mathcal{N}(0, 1)$, $w_{ij} \sim \mathcal{N}(0, 1/n)$, all independent.

CLT 3: An Application to Deep Learning



- Scenario: want to know behavior of randomly initialized one-layer neural net. Input $x = [x_1, \ldots, x_k]$ is dimension k, n hidden neurons, which each take value $\mathrm{ReLU}(\sum_{j=1}^{k} w_{ij} x_j)$. Output is number $f(x) = \sum_{i=1}^{n} a_i \mathrm{ReLU}(\sum_{j=1}^{k} w_{ij} x_j)$
- Neural net is randomly initialized with $a_i \sim \mathcal{N}(0, 1)$, $w_{ij} \sim \mathcal{N}(0, 1/n)$, all independent.
- Want to know distribution of $f(x)$ as number of hidden neurons $n \to \infty$, for FIXED x.

CLT 3: An Application to Deep Learning

- What is behavior of $\sum_{i=1}^{k} w_{ij} x_j$? Mean? Variance? Say $\|x\|_2^2 = \sum_{i=1}^{k} x_k^2 = 1$ for simplicity. Remember w_{ij} are independent and mean 0, $w_{ij} \sim \mathcal{N}(0, 1/n)$.

CLT 3: An Application to Deep Learning

- What is behavior of $\sum_{i=1}^{k} w_{ij} x_j$? Mean? Variance? Say $\|x\|_2^2 = \sum_{i=1}^{k} x_k^2 = 1$ for simplicity. Remember w_{ij} are independent and mean 0, $w_{ij} \sim \mathcal{N}(0, 1/n)$.
- Get $\sum_{i=1}^{k} w_{ij} x_j \sim \mathcal{N}(0, 1/n)$.

CLT 3: An Application to Deep Learning

- What is behavior of $\sum_{i=1}^{k} w_{ij} x_j$? Mean? Variance? Say $\|x\|_2^2 = \sum_{i=1}^{k} x_k^2 = 1$ for simplicity. Remember w_{ij} are independent and mean 0, $w_{ij} \sim \mathcal{N}(0, 1/n)$.
- Get $\sum_{i=1}^{k} w_{ij} x_j \sim \mathcal{N}(0, 1/n)$.
- What is behavior of $f(x) = \sum_{i=1}^{n} a_i \text{ReLU}(\sum_{j=1}^{k} w_{ij} x_j)$? Remember that $a_i \sim \mathcal{N}(0, 1)$ and ReLU(x) $= \max\{x, 0\}$ (x if positive, 0 otherwise). Take for fact that $\mathbb{E}\left[\text{ReLU}(\sum_{j=1}^{k} w_{ij} x_j)^2\right] \approx \frac{1}{2n}$ (look up "truncated normal distribution" if interested).

CLT 3: An Application to Deep Learning

- What is behavior of $\sum_{i=1}^{k} w_{ij} x_j$? Mean? Variance? Say $\|x\|_2^2 = \sum_{i=1}^{k} x_k^2 = 1$ for simplicity. Remember w_{ij} are independent and mean 0, $w_{ij} \sim \mathcal{N}(0, 1/n)$.
- Get $\sum_{i=1}^{k} w_{ij} x_j \sim \mathcal{N}(0, 1/n)$.
- What is behavior of $f(x) = \sum_{i=1}^{n} a_i \mathrm{ReLU}(\sum_{j=1}^{k} w_{ij} x_j)$? Remember that $a_i \sim \mathcal{N}(0, 1)$ and ReLU(x) $= \max\{x, 0\}$ (x if positive, 0 otherwise). Take for fact that $\mathbb{E}\left[\mathrm{ReLU}(\sum_{j=1}^{k} w_{ij} x_j)^2\right] \approx \frac{1}{2n}$ (look up "truncated normal distribution" if interested).
- Get $\mathbb{E}[f(x)] = 0$, $\mathrm{var}(f(x)) = \sum_{i=1}^{n} \mathbb{E}[a_i^2] \mathbb{E}\left[\mathrm{ReLU}(\sum_{j=1}^{k} w_{ij} x_j)^2\right] \approx 1/2$. Therefore, $f(x) \sim \mathcal{N}(0, 1/2)$ as $n \to \infty$ from CLT.

CLT 3: An Application to Deep Learning

- What is behavior of $\sum_{i=1}^{k} w_{ij}x_j$? Mean? Variance? Say $\|x\|_2^2 = \sum_{i=1}^{k} x_k^2 = 1$ for simplicity. Remember w_{ij} are independent and mean 0, $w_{ij} \sim \mathcal{N}(0, 1/n)$.
- Get $\sum_{i=1}^{k} w_{ij}x_j \sim \mathcal{N}(0, 1/n)$.
- What is behavior of $f(x) = \sum_{i=1}^{n} a_i \text{ReLU}(\sum_{j=1}^{k} w_{ij}x_j)$? Remember that $a_i \sim \mathcal{N}(0, 1)$ and $\text{ReLU(x)} = \max\{x, 0\}$ (x if positive, 0 otherwise). Take for fact that $\mathbb{E}\left[\text{ReLU}(\sum_{j=1}^{k} w_{ij}x_j)^2\right] \approx \frac{1}{2n}$ (look up "truncated normal distribution" if interested).
- Get $\mathbb{E}[f(x)] = 0$, $\text{var}(f(x)) = \sum_{i=1}^{n} \mathbb{E}[a_i^2]\mathbb{E}\left[\text{ReLU}(\sum_{j=1}^{k} w_{ij}x_j)^2\right] \approx 1/2$. Therefore, $f(x) \sim \mathcal{N}(0, 1/2)$ as $n \to \infty$ from CLT.
- Source: Prof. Hassani's topics class last semester.