Weak Law of Large Numbers & Central Limit Theorem

Math 30530, Spring 2019

April 30, 2019

Recall: (weak) Law of Large Numbers

$X_1, X_2, X_3, \ldots, X_n$ are independent copies of the same random variable, all with mean μ, variance σ^2.

$\bar{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$

Weak law of large numbers: For every degree of certainty $c < 1$, and every error margin $\varepsilon > 0$, there is an n large enough that

$$\Pr(|\bar{X}_n - \mu| \leq \varepsilon) \geq c.$$

Interpretation: Repeat an experiment many times independently, record the answers, and average them. By making "many" large enough, we can be sure that the average of the readings is arbitrarily close to the theoretical experiment average, with arbitrarily high probability (arbitrarily close to 1)

Effective version

$X_1, X_2, X_3, \ldots, X_n$ are independent copies of the same random variable, all with mean μ, variance σ^2, $\bar{X}_n = \frac{X_1+X_2+\ldots+X_n}{n}$

Weak law of large numbers: For every k,

$$\Pr\left(|\bar{X}_n - \mu| \leq \frac{k\sigma}{\sqrt{n}}\right) \geq 1 - \frac{1}{k^2}$$

Example: A certain brand of lightbulb has lifetime that is exponentially distributed with mean A hours, A unknown. I try to estimate A by letting n lightbulbs run independently, and recording & averaging their lifetimes. How large should n be, so that I can be at least 90% sure that the estimate I get is within 5% of the actual average A?

Example — lifetime of lightbulb

Example: A certain brand of lightbulb has lifetime that is exponentially distributed with mean A hours, A unknown. I try to estimate A by letting n lightbulbs run independently, and recording & averaging their lifetimes. How large should n be, so that I can be at least 90% sure that the estimate I get is within 5% of the actual average A?

Let X_i be lifetime of ith lightbulb. Have $X_i \sim \mathrm{Exponential}(\lambda)$ (some unknown λ), $\mu = 1/\lambda = A$, $\sigma^2 = 1/\lambda^2 = A^2$.

With $\bar{X}_n = (X_1 + \ldots + X_n)/n$, using $\Pr(|\bar{X}_n - \mu| \leq k\sigma/\sqrt{n}) \geq 1 - 1/k^2$, get

$$\Pr\left(|\bar{X}_n - A| \leq \frac{kA}{\sqrt{n}}\right) \geq 1 - \frac{1}{k^2}.$$

Want $1 - 1/k^2 = 0.9$ so $k = \sqrt{10}$, then want $k/\sqrt{n} = \sqrt{10/n} = 0.05$, so $n = 4000$.

Example — accumulating rounding error

Example: I estimate the sum of n random real numbers by rounding each to the nearest integer, and adding the resulting integers. What is the error?

Let X_i be error in ith number, so $X_i \sim \mathrm{Uniform}(-1/2, 1/2)$. Let T_n be the total error, so $T_n = X_1 + \ldots + X_n$, all X_i independent. Each X_i has $\mu = 0$, $\sigma^2 = 1/12$. By weak law of large numbers,

$$\Pr\left(|\bar{X}_n - 0| \leq \frac{k}{\sqrt{12n}}\right) \geq 1 - \frac{1}{k^2},$$

so, since $T_n = n\bar{X}_n$,

$$\Pr\left(|T_n| \geq \frac{k\sqrt{n}}{\sqrt{12}}\right) \leq \frac{1}{k^2}.$$

E.g., when $n = 1000$, to find $P(|T_{1000}| > 25)$ solve $k\sqrt{1000/12} = 25$, so $k \approx 2.738$, so

$$P(|T_{1000}| > 25) \leq 0.133\cdots.$$

Central Limit Theorem

Informally: Add together lots of independent copies of a random variable. The result is very close to being a normal random variable.

Less informally: Let $X_1, X_2, \ldots, X_n$ be a collection of independent copies of the same random variable, with X_i having mean μ and variance σ^2. Then

$$X_1 + \ldots + X_n \approx \mathrm{Normal}(n\mu, n\sigma^2).$$

Precisely: Let $X_1, X_2, \ldots, X_n$ be as above. Set

$$S_n = \frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n}\sigma}$$

(so $E(S_n) = 0$, $\mathrm{Var}(S_n) = 1$). Then, for each $x \in \mathbb{R}$,

$$\Pr(S_n \leq x) \to \Pr(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2}\, dz$$

as $n \to \infty$.

Example — lifetime of lightbulb

Example: A certain brand of lightbulb has lifetime that is exponentially distributed with mean A hours, A unknown. I try to estimate A by letting n lightbulbs run independently, and recording & averaging their lifetimes. How large should n be, so that I can be at least 90% sure that the estimate I get is within 5% of the actual average A?

Let X_i be lifetime of ith lightbulb. Have $X_i \sim \mathrm{Exponential}(\lambda)$ (some unknown λ), $\mu = 1/\lambda = A$, $\sigma^2 = 1/\lambda^2 = A^2$.

So $X_1 + \cdots + X_n \sim \mathrm{Normal}(nA, nA^2)$, and $\bar{X}_n - A \sim \mathrm{Normal}(0, A^2/n)$.

$$\begin{aligned} P(|\bar{X}_n - A| \leq 0.05A) &\approx P(|Z| \leq 0.05\sqrt{n}) \\ &\approx 0.9 \text{ when } 0.05\sqrt{n} \approx 1.645 \end{aligned}$$

So take $n \approx 1082$ (much better than $n = 4000$ from Law of Large Numbers)

Example — accumulating rounding error

Example: I estimate the sum of n random real numbers by rounding each to the nearest integer, and adding the resulting integers. What is the error?

Let X_i be error in ith number, so $X_i \sim \mathrm{Uniform}(-1/2, 1/2)$. Let T_n be the total error, so $T_n = X_1 + \ldots + X_n$, all X_i independent. Each X_i has $\mu = 0$, $\sigma^2 = 1/12$. By central limit theorem, $T_n \approx \mathrm{Normal}(0, n/12)$, so

$$\begin{aligned}\Pr(|T_n| \le t) &\approx \Pr(-t \le \mathrm{Normal}(0, n/12) \le t) \\ &= \Pr\left(-\frac{\sqrt{12}t}{\sqrt{n}} \le Z \le \frac{\sqrt{12}t}{\sqrt{n}}\right)\end{aligned}$$

Example: when $n = 1000$ and $t = 25$,

$$P(|T_{1000}| > 25) \approx P(|Z| > 2.738) \approx 0.0062$$

(much better than ≤ 0.133 from Law of Large Numbers)

Hint of proof of CLT

$X_1, X_2, \ldots, X_n$ a collection of independent copies of the same random variable, with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Set

$$S_n = \frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n}\sigma}$$

(so $E(S_n) = 0$, $\mathrm{Var}(S_n) = 1$). Want $S_n \to Z$ (standard normal).

$$S_n = \frac{X_1 - \mu}{\sqrt{n}\sigma} + \cdots + \frac{X_n - \mu}{\sqrt{n}\sigma} = Y_1 + \cdots + Y_n.$$

$E(Y_i) = 0$, $\mathrm{Var}(Y_i) = 1/n$ so $E(Y_i^2) = 1/n$, so

$$M_{Y_i}(t) = 1 + 0t + \frac{t^2}{2n} + \text{terms in } t^3, t^4 \text{ et cetera.}$$

For small t, $M_{S_n}(t) = M_{Y_1}(t) \cdots M_{Y_n}(t) \approx \left(1 + \frac{t^2}{2n}\right)^n \to e^{\frac{t^2}{2}} = M_Z(t)$. So

$$S_n \to Z \text{ as } n \to \infty$$