Contents

Distribution of Sample Mean

Consider N random variables $X_1, \dots, X_N$

- Let's consider $\bar{X} = \frac{1}{N} \sum_{k=1}^{N} X_k$
- $\bar{X}$ is called the "sample mean" or the "empirical mean"
- $\bar{X}$ is a random variable

Suppose we observe values for $X_1, \dots, X_N$ and calculate the empirical mean of the observed values. That gives us one value for $\bar{X}$. But the value of $\bar{X}$ changes depending on the observed values.

- Suppose we toss a fair coin $N = 5$ times and get H, H, H, T, T. Let $X_k = 1$ when comes up heads. Then $\frac{1}{N} \sum_{k=1}^{N} X_k = \frac{3}{5}$
- Suppose we toss the coin another $N = 5$ times and get T, T, H, T, H. Now $\frac{1}{N} \sum_{k=1}^{N} X_k = \frac{2}{5}$

Random variable $\bar{X} = \frac{1}{N} = \sum_{k=1}^{N} X_k$

- Suppose the X_k are **independent and identically distributed**
- Each X_k has mean $E[X_k] = \mu$ and variable $Var(X_k) = \sigma^2$

Then we can calculate the mean of $\bar{X}$ as

$$E[\bar{X}] = E(\frac{1}{N} \sum_{k=1}^{N} X_k) = \frac{1}{N} \sum_{k=1}^{N} E[X_k] = \mu$$

BK: recall linearity of expectation: $E[X+Y] = E[X]+E[Y]$ and $E[aX] = aE[X]$

- We say $\bar{X}$ is an **unbiased estimator** or amu since $E[\bar{X}] = x$

We can calculate the variance of $\bar{X}$ as

$$Var(\bar{X}) = Var(\frac{1}{N}\sum_{k=1}^{N} X_k) = \frac{1}{N^2}Var(\sum_{k=1}^{N} X_k) = \frac{1}{N^2}\sum_{k=1}^{N} Var(X_k) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$$

NB: recall $Var(aX) = a^2Var(X)$ and $Var(X + Y) = Var(X) + Var(Y)$ when X, Y are independent.

- As N increases, the variance of $\bar{X}$ falls
- $Var(NX) = N^2Var(X)$ for random variable X
- But when add together **indepdent** random variable $X_1 + X_2 + \ldots$ the variance is only $NVar(X)$ rather than $N^2Var(X)$
- This is due to **statistical multiplexing**
  - Small and large values of X_i tend to cancel out for large N

Weak Law of Large Numbers

Consider N independent and identically distributed random variables $X_1, \ldots, X_N$ each with mean μ and variance σ^2. Let $\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k$. For any $\epsilon > 0$:

$$P(\mid \bar{X} - \mu \mid \geq \epsilon) \to 0 \text{ as } N \to \infty$$

That is, $\bar{X}$ **concentrates** around the mean μ as N increases.

Who Cares?

- Suppose we have an event E
- Define indicator random variable X_i equal to 1 when event E is observed in trial i and 0 otherwise
- Recall $E[X_I] = P(E)$ is the probability that event E occurs
- $\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k$ is then the relative frequency with which event E is observed over N experiments
- And $P(\mid \bar{X} - \mu \mid \geq \epsilon) \to 0$ as $N \to \infty$ tells us that this observed relative frequency $\bar{X}$ converges to the probability $P(E)$ of event E as N grows large
- So the law of large numbers formalises the intuition of probability as frequency when an experiment can be repeated many times.
  - But probability still makes sense even if cannot repeat an experiment many times - all our analysis still holds

Confidence Intervals

- Recall that when a random variable lies in an interbal $a \leq X \leq b$ with a specified probability we call this a confidence interval, e.g. $p - 0.05 \leq T \leq p + 005$ with probability at least 0.95
- Chebyshev inequality allows us to calculate confidence intervals given the mean and variance of a random variable
- For sample mean $\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k$, Chebyshev inequality tells us $P(\mid \bar{X} - \mu \mid \geq \epsilon) \leq \frac{\sigma^2}{N\epsilon^2}$ when μ is mean of X_k and σ^2 is its variance
- E.g. When $\epsilon = \frac{\sigma}{\sqrt{0.05N}}$ then $\frac{\sigma^2}{N\epsilon^2} = 0.05$ and Chebyshev tells us that $\mu - \frac{\sigma}{\sqrt{0.05N}} \leq \bar{X} \leq \mu + \frac{\sigma}{\sqrt{0.05N}}$ with probability at least 0.95

Bootstrapping

Mean μ, variance σ^2 and number of points N summarises our N data points using three numbers. But we have $N \gg 3$ data points. Can we use these to also empirically estimate the *distribution$ of the sample mean? Yes!

- Make use of the fact that computing power is cheap
- The N data points are drawn independently from the same probability distribution F
- So the idea is to use these N data points as a surrogate for F
  - To generate new samples from F we draw uniformly at random from our N data points
  - This is **sampling with replacement**
- Suppose our data if $\{A, B, C, D, E, F\}$
  - Select one point uniformly as random, e.g. B
  - Select a second point uniformly at random, might be B again or might be something else
  - And so on until we desired number of samples
- Bootstrapping is an example of a **resampling method**

Bootstrapping:

1. Draw a sample of N data points uniformly at random from data, with replacement
2. Using this sample estimate the mean $X_1 = \frac{1}{N}\sum_{i=1}^{N} X_{1,i}$
3. Repeat, to generate a set of estimates $X_1, X_2, \ldots$

- The distribution of these estimates approximates the distribution of the sample mean (it is not exact)

Example: coin tossing!

- Toss $N = 100$ biased coins, lands heads with probability $p = 0.1$
  - $X_i = 1$ if ith toss is heads, 0 otherwise
- Sample with replacement from the $N = 100$ data points
- Calculate sample mean $X = \frac{1}{N}\sum_{i=1}^{N} X_i$
- Repeat 1000 times and plot observed distribution of X

Note: bootstrap estimate of the disribution is only approximate

- Different data leads to different estimates of the distribution
- But very handy all the same
- Using our empirical estimate of the distribution of the sample mean X we can estimate confidence intervals, etc.