The Markov Inequality

- Suppose x is a non-negative random variable
- Markov inequality upper bounds the probability that x is large

$$\mathbf{P}\left(x \geq a\right) \leq \frac{E[x]}{a} \tag{1}$$

- To use it only information that is needed is the mean of the r.v.
- No need for the P.M.F/P.D.F of the r.v.

The Chebyshev Inequality

- Chebyshev inequality is:

$$\mathbf{P}\left(|x - E[x]| \geq t\right) \leq \frac{var(x)}{t^2}$$

- Only need mean and the variance of the r.v.
- Proof: Apply the Markov inequality to the new r.v $|x - E[x]|^2$
- The Chebyshev inequality can also be written as:

$$\mathbf{P}\left(x \in [E[x] - t, E[x] + t]\right) \geq 1 - \frac{var(x)}{t^2} \tag{2}$$

- The interval $[E[x] - t, E[x] + t]$ is called confidence interval
- $1 - \frac{var(x)}{t^2}$ is called confidence level

Examples of Markov/Chebyshev inequality

- A post office handles, on average, 10,000 letters a day. What can be said about the probability that it will handle at least 15,000 letters tomorrow.
- A post-office handles 10,000 letters per day with a variance of 2,000 letters. What can be said about the probability that this post office handles between 8,000 and 12,000 letters tomorrow? What about the probability that more than 15,000 letters come in?
- A faulty coin with probability of head as 0.2 is tossed 10 times. What is the bound on the probability that 8 heads show up?

Sequence, Limits

- Sequence a_n, limit number a
- a_n converges to a

$$\lim_{n \to \infty} a_n = a \tag{3}$$

- "a_n eventually gets and stays arbitrarily close to a"
- For every $\epsilon > 0$, there exists n_0 such that for every $n > n_0$, we have

$$|a_n - a_0| \leq \epsilon$$

Convergence in Probability

- Sequence of random variables Y_n
- Converges in probability to a number a
- "Almost all of the PDF of Y_n eventually gets concentrated close to a"
- For every $\epsilon > 0$

$$\lim_{n \to \infty} \mathbf{P}\left(|Y_n - a|\right) \geq \epsilon = 0$$

- Example: Does Y_n converge?
  $\mathbf{P}\left(Y_n = 0\right) = 1 - \frac{1}{n} \quad \mathbf{P}\left(Y_n = 1\right) = \frac{1}{n}$

Convergence of the sample mean: WLLN

- $X_1, X_2 \dots$ are i.i.d
- finite mean μ and variance σ^2
- $M_n = \frac{X_1+X_2+\cdots+X_n}{n}$
- $E[M_n] =$
- $var(M_n) =$

$$\mathbf{P}\left(|M_n - \mu| \geq \epsilon\right) \leq \frac{var(M_n)}{\epsilon^2} = \frac{\sigma^2}{n\,\epsilon^2}$$

- M_n converges in probability to μ

Polling Problem

- f : fraction of population that ".."
- i^{th} person polled: $X_i = 1$ if yes, $X_i = 0$ if no.
- M_n is thus the fraction of yes in our sample.
- Goal: 95% confidence and less than 1% error

$$\mathbf{P}\left(|M_n - f| \geq 0.01\right) \leq 0.05$$

- Use Chebyshev inequality:

$$\mathbf{P}\left(|M_n - f| \geq 0.01\right) \leq \frac{\sigma_x^2}{n\,0.01^2} \leq \frac{1}{4\,n\,0.01^2}$$

- $n = 50,000$. But this is conservative

Different Scalings of M_n

- $X_1, \ldots, X_n$ i.i.d
- Three variants of their sum
- $S_n = X_1 + \ldots X_n$, variance $n\sigma^2$
- $M_n = \frac{S_n}{n}$, variance $\frac{\sigma^2}{n}$: WLLN
- $\frac{S_n}{\sqrt{n}}$: constant variance σ^2. Asymptotic shape?

The Central Limit Theorem

- Standardized $S_n = X_1 + \ldots X_n$:

$$Z_n = \frac{S_n - E[S_n]}{\sqrt{n}\sigma}$$

- Zero mean, Unit Variance
- Let Z be a standard normal r.v.
- Theorem: For every c:

$$\mathbf{P}(Z_n \leq c) \to \mathbf{P}(Z \leq c)$$

C.L.T

- Treat Z_n, therefore S_n as if normal
- universal; only means, variances matter
- accurate computational shortcut
- Can be used even for moderate n
- Used in every walk of life.

Polling Problem: Using C.L.T

- f : fraction of population that ".."
- i^{th} person polled: $X_i = 1$ if yes, $X_i = 0$ if no.
- M_n is thus the fraction of yes in our sample.
- Goal: 95% confidence and less than 1% error

$$\mathbf{P}\left(|M_n - f| \geq 0.01\right) \leq 0.05$$

- Event of interest:

$$\mathbf{P}\left(|M_n - f| \geq 0.01\right)$$

- Using C.L.T

$$\mathbf{P}\left(|M_n - f|\right) > 0.01 \approx \mathbf{P}\left(|Z| > 0.01\frac{\sqrt{n}}{\sigma}\right)$$

Applying C.L.T to Binomal Distribution

- Let X_i be Bernoulli(p)
- $S_n = X_1 + X_2 \dots X_n$, Binomial r.v.
- mean $= np$, variance $= np(1-p)$
- CDF of $\frac{S_n - np}{\sqrt{np(1-p)}} \rightarrow$ standard normal
- Example: $n = 36$, $p = 0.5$, $\mathbf{P}\left(S_n \leq 21\right)$
- $\sum_0^{21} \binom{36}{i} (0.5)^{36} = 0.8785$

The $\frac{1}{2}$ correction for binomial approximation

- $\mathbf{P}(S_n \leq 21) = \mathbf{P}(S_n < 22)$
- Consider $S_n < 21.5$
- De Moivre - Laplace CLT (for binomial):
- When the 1/2 correction is used, CLT can also approximate the binomial p.m.f. (not just the binomial CDF)
- $\mathbf{P}(S_n = 19) = \mathbf{P}(18.5 \leq S_n \leq 19.5)$
- $\frac{18.5-18}{3} \leq \frac{S_n-18}{3} \leq \frac{19.5-18}{3}$
- $0.17 \leq Z_n \leq 0.5$
- $\mathbf{P}(Z \leq 0.5) - \mathbf{P}(Z \leq 0.17) = 0.124.$
- Exact answer: 0.1251