Contents

Question: toss a fair coin $n = 10000$ times. How many heads?

$X = \sum_{i=1}^{n}$, $X_i \sim \text{Bern}(1/2)$. $\mathbb{E}[X] = 5000$. But $\mathbb{P}(X = 5000) = \binom{10^4}{5000} \cdot 2^{-10^4} \approx 0.008$.

Theorem 0.1 (Weak Law of Large Numbers) Let $X_1, ..., X_n$ be IID RVs with mean $\mathbb{E}[X_1] = \mu$. Then, for all $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\right) \to 0 \quad \text{as } n \to \infty.$$

So $\mathbb{P}(X \in [5000 - n\varepsilon, 5000 + n\varepsilon]) \approx 1$.

Theorem 0.2 (Central Limit Theorem) Let $X_1, ..., X_n$ be IID RVs with mean $\mathbb{E}[X_1] = \mu$. Let $\text{Var}(X_1) = \sigma^2 < \infty$. Then $\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n}(X_i - \mu) \underset{D}{\to} N(0, 1)$, i.e.

$$\mathbb{P}\left(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) \in A\right) \to \int_A \frac{1}{\sqrt{2n}} e^{-x^2/2} \, \mathrm{d}x$$

for all A.

So $\mathbb{P}\left(X \in \left[\frac{n}{2} - \frac{\sqrt{n}}{2}Q^{-1}(\delta), \frac{n}{2} + \frac{\sqrt{n}}{2}Q^{-1}(\delta)\right]\right) \geq 1 - \delta$, for n large enough, where $Q(\delta) = \int_{\delta}^{\infty} \frac{1}{\sqrt{2n}} e^{-x^2/2d} \, \mathrm{d}x$. We have $Q^{-1}(x) \propto \sqrt{\log \frac{1}{x}}$. So interval has length $\propto \sqrt{n}\sqrt{\log \frac{1}{\delta}}$.

Theorem 0.3 (Chebyshev's Inequality) $\mathbb{P}(|X - \mu| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}$ for all $\varepsilon > 0$.

Corollary 0.4 $\mathbb{P}\left(\left|\sum_{i=1}^{n}(X_i) - \frac{n}{2}\right| \geq t\right) \leq \frac{\text{Var}\left(\sum_{i=1}^{n} X_i\right)}{t^2} = n\frac{\sigma^2}{t^2} \leq \delta$ where $t = \sqrt{n}\sigma/\sqrt{\delta}$.

So $\mathbb{P}(X \in [\frac{n}{2} -, \frac{n}{2}]) \geq 1 - \delta$.

Question 2: we have N coupons. Each day receive one uniformly at random independent of the past. How many days until all coupons received?

We have $X = \sum_{i=1}^{n} X_i$, $X_i \sim \text{Geom}\left(\frac{i}{n}\right)$. $\mathbb{E}[X] = \sum_i \mathbb{E}[X_i] \approx n \log n$ (verify this).

Question 3: Let $(X_1, ..., X_n), (Y_1, ..., Y_n)$ be IID. What is the longest common subsequence, i.e. $f(X_1, ..., X_n, Y_1, ..., Y_n) = \max\left\{k : \exists i_1, ..., i_k, j_1, ..., j_k \text{ s.t. } X_{i_j} = Y_{i_j} \; \forall j \in [k]\right\}$. Computing f is NP-hard. f is smooth.

Principle: a smooth function of many independent random variables concentrates around its mean.

Tower property of conditional expectation: $\mathbb{E}(\mathbb{E}(Z \mid X, Y) \mid Y) = \mathbb{E}(Z \mid Y)$.

1. The Chernoff-Cramer method

Theorem 1.1 (Markov's Inequality) Let Y be a non-negative RV. For any $t \geq 0$,

$$\mathbb{P}(Y \geq t) \leq \frac{\mathbb{E}[Y]}{t}.$$

Proof. We have $Y = Y\mathbb{I}_{\{Y \geq t\}} + Y\mathbb{I}_{\{Y < t\}} \geq t \cdot \mathbb{I}_{\{Y \geq t\}}$. Taking expectations gives the result. □

Corollary 1.2 (Chebyshev's Inequality) $\mathbb{P}(|Y-\mathbb{E}[Y]| \geq t) \leq \frac{\mathbb{E}[Y-\mathbb{E}[Y]]^2}{t^2}$

Proof. Take $Z = Y - \mathbb{E}[Y]$ and use Markov's. $\square$

Let $\varphi : \mathbb{R} \to \mathbb{R}_+$ be non-decreasing, then $\mathbb{P}(\varphi(Y) \geq \varphi(t)) \leq \frac{\mathbb{E}[\varphi(Y)]}{\varphi(t)}$. For $\varphi(t) = t^2$, we can use $\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i)$.

Exercise 1.3 Prove WLLN, assuming that $\operatorname{Var}(X_1) < \infty$, using Chebyshev's inequality.

Notation 1.4 For $\lambda > 0$, let $\varphi_\lambda(t) = e^{\lambda t}$.

Note $\mathbb{P}(Z \geq t) = \mathbb{P}\left(e^{\lambda Z} \geq t\right) \leq \frac{\mathbb{E}\left[e^{\lambda Z}\right]}{\varphi_\lambda(t)} = e^{-(\lambda t - \log(\mathbb{E}[e^{\lambda Z}]))}$. So if $\mathbb{E}\left[e^{\lambda Z}\right] < \infty$, then we have exponential concentration.

$F(\lambda) := \mathbb{E}\left[e^{\lambda Z}\right] = \sum_{i=0}^\infty \frac{\lambda^i \mathbb{E}[Z^i]}{i!}$. We have $\varphi_Z(\lambda) = \log(F(\lambda))$ is additive: if $Z = \sum_{i=1}^n Z_i$, Z_i independent, then $\varphi_Z(\lambda) = \log\left(\mathbb{E}\left[e^{\lambda Z}\right]\right) = \sum_i \log \mathbb{E}\left[e^{\lambda Z_i}\right]$.

So $\mathbb{P}(Z \geq t) \leq \inf_{\lambda > 0} e^{-(\lambda t - \varphi_\lambda(Z))} = e^{-\sup_{\lambda>0}(\lambda t - \varphi_Z(\lambda))}$. This is the Chernoff bound. We denote $\varphi_Z^*(t) = \sup_{\lambda > 0} \lambda t - \varphi_Z(\lambda)$. This is Cramer's transform of Z.

Goal is to obtain upper bound on $\varphi_Z(\lambda)$, as this will give concentration. The function $\varphi_{Z-\mathbb{E}[Z]}(\lambda)$ gives upper bounds on $\mathbb{P}(Z - \mathbb{E}[Z] \geq t)$, the function $\varphi_{-Z+\mathbb{E}[Z]}(\lambda)$ gives upper bounds on $\mathbb{P}(Z - \mathbb{E}[Z] \leq -t)$.

Proposition 1.5 Properties of $\varphi_Z(\lambda)$:

1. $\varphi_Z(\lambda)$ is convex and infinitely differentiable on (a, b), where $b = \sup_{\lambda>0}\left\{\mathbb{E}\left[e^{\lambda Z}\right] < \infty\right\}$.
2. $\varphi_Z^*(t) \geq 0$ and convex.
3. If $t > \mathbb{E}[Z]$, then $\varphi_Z^*(t) = \sup_{\lambda \in \mathbb{R}}(\lambda t - \varphi_Z(\lambda))$, the Fenchel-Legendre dual.