

Question 1: You toss a coin 10,000 times. How many heads do you see?

Question 2: Coupon collector problem. Have N coupons and we need to collect them all. How many coupons do we need to sample to get all N ?

Question 3: Largest common subsequence problem: have sequences X_1, \dots, X_n and Y_1, \dots, Y_n of iid Bern(1/2) random variables. What is the largest k such that there exist $i_1 < i_2 < \dots < i_k$ and $j_1 < j_2 < \dots < j_k$ such that $X_{i_1} = Y_{j_1}, \dots, X_{i_k} = Y_{j_k}$?

Question 1: we have various possible answers:

- 5,000. Indeed if we let X_i be the indicator of the event that we see heads on the i th toss, the number of heads is $S = \sum_{i=1}^{10000} X_i$ and $\mathbb{E}S = 5000$. But $\mathbb{P}(S = 5000) = \binom{10000}{5000} 2^{-10000} \approx 0.008$.
- Weak Law of Large Numbers: let $(X_i)_{i \geq 1}$ be iid with finite expectation μ and finite second moments. Then for every $\varepsilon > 0$,

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

Therefore for large enough n , the number of heads lies in $[n(1/2 - \varepsilon), n(1/2 + \varepsilon)]$ with high probability. The main problem is that this is an asymptotic result - we don't know how large n should be.

- Central Limit Theorem: let $(X_i)_{i \geq 1}$ be iid with finite mean μ and finite second moment $\sigma^2 + \mu^2$. Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \xrightarrow{d} \mathcal{N}(0, 1).$$

Therefore $\sum_{i=1}^n (X_i - \mu)$ has deviations of the order $\sqrt{n}\sigma$. Suppose we pretend 10000 is big: then

$$\begin{aligned} S = \sum_{i=1}^{10000} X_i &\in [5000 - Q^{-1}(0.005)\sqrt{100}/2, 5000 + Q^{-1}(0.005)\sqrt{100}/2] \\ &\approx [5000 \pm 128] \end{aligned}$$

with probability 0.99, where $Q(x) = \mathbb{P}(Z \geq x)$ for $Z \sim \mathcal{N}(0, 1)$. However we have the same issue again - is $n = 10000$ large enough?

We can however give some non-asymptotic answers to Question 1:

Proposition (Chebyshev's inequality). Let X be any random variable with mean μ and variance σ^2 . Then

$$\mathbb{P}(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}.$$

With this, we have

$$\mathbb{P}\left(\left|\sum_{i=1}^{10000} X_i - 5000\right| > t\right) \leq \frac{10000 \times \frac{1}{4}}{t^2} = \frac{2500}{t^2}.$$

So in particular, if $t = 500$ the RHS is 0.01. So we have $S \in [4500, 5500]$ with probability 0.99. However note that this is a weaker result than what the Central Limit Theorem gives.

Question 2: the number of samples S is equal to $\sum_{i=1}^N X_i$ where $X_i \sim \text{Geo}(i/N)$. Thus $\mathbb{E}S = \sum_{i=1}^N \frac{N}{i} = N \sum_{i=1}^N \frac{1}{i} \approx N \log N$.

Question 3: we have a function $f(X_1, \dots, X_n, Y_1, \dots, Y_n)$ which gives the longest common subsequence. It turns out this function is “smooth” in a certain sense, for which we can use “Talagrand’s Principle”.

Chernoff-Cr  mer method

Theorem (Markov's inequality). *Let Y be a non-negative random variable with finite expectation. Then for any $t > 0$ we have*

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

Proof. Note $t\mathbb{1}(Y \geq t) \leq Y$ and integrate. \square

Corollary. *Let Y be a random variable. Suppose $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ is increasing and such that $\mathbb{E}|\phi(Y)| < \infty$. Then*

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

Note that for a random variable Z , letting $Y = |Z - \mathbb{E}Z|$ and $\phi : t \mapsto t^2$ gives Chebyshev's inequality $\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq \frac{\text{Var}(Z)}{t^2}$.

Could also take $\phi : t \mapsto t^q$ for any $q > 0$ to conclude $\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq \frac{\mathbb{E}|Z - \mathbb{E}|^q}{t^q}$.

Consider instead $\phi : t \mapsto e^{\lambda t}$ for $\lambda > 0$. Then we get

$$\mathbb{P}(Z \geq t) \leq \frac{\mathbb{E}[e^{\lambda Z}]}{e^{\lambda t}}.$$

Define $F(\lambda) = \mathbb{E}e^{\lambda Z}$, the *moment generating function* of Z . Define $\psi_Z(\lambda) = \log \mathbb{E}[e^{\lambda Z}]$. If X_1, \dots, X_n are independent and $Z = \sum_{i=1}^n X_i$ then it is clear that $\psi_Z(\lambda) = \sum_{i=1}^n \psi_{X_i}(\lambda)$. So we have

$$\mathbb{P}(Z \geq t) \leq \inf_{\lambda \geq 0} e^{\psi_Z(\lambda) - \lambda t}.$$

Now define $\psi_Z^*(t) = \sup_{\lambda \geq 0} (\lambda t - \psi_Z(\lambda))$ and write $\mathbb{P}(Z \geq t) \leq e^{-\psi_Z^*(t)}$. This is known as the *Chernoff bound*, and ψ_Z^* is known as the *Chernoff-Cr  mer transform*.

Properties of ψ_Z and ψ_Z^*

1. ψ_Z is convex and infinitely differentiable on $(0, b)$ where $b = \sup\{\lambda : \psi_Z(\lambda) < \infty\}$. Indeed

$$\begin{aligned} F(\theta x + (1 - \theta)y) &= \mathbb{E}[e^{\theta x Z} e^{(1 - \theta)y Z}] \\ &\leq \mathbb{E}[e^{x Z}]^\theta \mathbb{E}[e^{y Z}]^{1 - \theta}. \end{aligned} \quad (\text{H  lder with } 1/p = \theta, 1/q = 1 - \theta)$$

2. $\psi_Z^* \geq 0$ and it is convex (follows from the definition).

3. Suppose $t \geq \mathbb{E}Z$. Then $\psi_Z^*(t) = \sup_{\lambda} (\lambda t - \psi_Z(\lambda))$. Indeed we'll show $\lambda t - \psi_Z(\lambda) \leq 0$ whenever $\lambda < 0$. We have

$$\begin{aligned}\mathbb{E}[e^{\lambda Z}] &\geq e^{\lambda \mathbb{E}Z} && \text{(Jensen)} \\ \implies \psi_Z(\lambda) &\geq \lambda \mathbb{E}Z \\ \implies \lambda t - \psi_Z(\lambda) &\leq \lambda t - \lambda \mathbb{E}Z = \lambda(t - \mathbb{E}Z) \leq 0.\end{aligned}$$

Example. Let $Z \sim \mathcal{N}(0, v)$. We want to upper bound $\mathbb{P}(Z \geq t)$ for $t > 0$. We have

$$\begin{aligned}\mathbb{E}[e^{\lambda Z}] &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi v}} e^{-\frac{t^2}{2v}} e^{\lambda t} dt \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi v}} e^{-\frac{(t-\lambda v)^2}{2v}} e^{\frac{v\lambda^2}{2}} dt \\ &= e^{\frac{v\lambda^2}{2}}.\end{aligned}$$

Hence $\psi_Z^*(t) = \sup_{\lambda} \left(\lambda t - \frac{\lambda^2 v}{2} \right)$ (for $t > 0 = \mathbb{E}Z$). Differentiating we see the optimal value is $\lambda = t/v$. Plugging this in gives $\psi_Z^*(t) = \frac{t^2}{2v}$. Thus

$$\mathbb{P}(Z \leq t) \leq e^{-\frac{t^2}{2v}}.$$

Sub-Gaussian random variables

Definition. A random variable Y with $\mathbb{E}Y = 0$ is *sub-Gaussian* with variance parameter v if

$$\psi_Y(\lambda) < \frac{\lambda^2 v}{2} \quad \forall \lambda \in \mathbb{R}.$$

The set of sub-Gaussian random variables with variance parameter v is denoted $\mathcal{G}(v)$.

1. It is clear from the above that if $Y \in \mathcal{G}(v)$ then $\mathbb{P}(Y \geq t) \leq e^{-t^2/2v}$ and $\mathbb{P}(Y \leq -t) \leq e^{-t^2/2v}$.
2. If $Y_i \in \mathcal{G}(v_i)$ for $i = 1, \dots, n$ are independent then $\sum_{i=1}^n Y_i \in \mathcal{G}(\sum_{i=1}^n v_i)$ (immediate by additivity of $\psi(\cdot)$).
3. If $Y \in \mathcal{G}(v)$ then $\text{Var}(Y) \leq v$ (see Example Sheet).

Theorem. The following are equivalent for suitable v, b, c, d

1. $Y \in \mathcal{G}(v)$;
2. $\max\{\mathbb{P}(Y \geq t), \mathbb{P}(Y \leq -t)\} \leq e^{-\frac{t^2}{2b}}$ for all $t > 0$;
3. $\mathbb{E}Y^{2q} \leq q!c^q$ for all $q \geq 1$;
4. $\mathbb{E}[e^{dY^2}] \leq 2$.

Proof. Not given. □

Lemma (Hoeffding's lemma). Let Y be supported on $[a, b]$ and suppose $\mathbb{E}Y = 0$. Then $\psi_Y''(\lambda) \leq \frac{(b-a)^2}{4}$, and so $Y \in \mathcal{G}\left(\frac{(b-a)^2}{4}\right)$.

Proof. We have

$$\psi'_Y(\lambda) = \frac{\mathbb{E}[Y e^{\lambda Y}]}{\mathbb{E}[e^{\lambda Y}]} \implies \psi''_Y(\lambda) = \frac{\mathbb{E}[e^{\lambda Y}] \mathbb{E}[Y^2 e^{\lambda Y}] - (\mathbb{E}[Y e^{\lambda Y}])^2}{\mathbb{E}[e^{\lambda Y}]^2}.$$

So

$$\begin{aligned} \psi''_Y(\lambda) &= \int_{\mathbb{R}} y^2 \underbrace{\frac{e^{\lambda y}}{\mathbb{E}[e^{\lambda Y}]}}_{:=dQ(y)} d\mu_Y(y) - \left(\int_{\mathbb{R}} y \frac{e^{\lambda y}}{\mathbb{E}[e^{\lambda Y}]} d\mu_Y(y) \right)^2 \\ &= \text{Var}_{Y \sim Q}(Y) \geq 0 \end{aligned}$$

noting that Q is supported on $[a, b]$. If $Y \in [a, b]$ almost-surely then note

$$\text{Var}(Y) = \text{Var}\left(Y - \frac{a+b}{2}\right) \leq \mathbb{E}\left[\left(Y - \frac{a+b}{2}\right)^2\right] \leq \frac{(b-a)^2}{4}.$$

To finish, observe that $\psi_Y(\lambda) = \psi_Y(0) + \lambda \psi'_Y(0) + \frac{\lambda^2}{2} \psi''_Y(\theta)$ for some $\theta \in [0, \lambda]$. Thus $\psi_Y(\lambda) \leq \frac{\lambda^2(b-a)^2}{8}$. \square

Theorem (Hoeffding's inequality). *Let Y_1, \dots, Y_n be independent random variables with Y_i having support on $[a_i, b_i]$. Then*

$$\mathbb{P}\left(\sum_{i=1}^n (Y_i - \mathbb{E}Y_i) \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Proof. Trivial by Hoeffding's lemma and additivity of the variance parameters. \square

Theorem (Bennett's inequality). *For $1 \leq i \leq n$, let X_i be independent random variables satisfying $\mathbb{E}X_i = 0$, $\text{Var}(X_i) = \sigma_i^2$ and let $v = \sum_{i=1}^n \sigma_i^2$. Further assume the X_i are bounded by some $C > 0$ almost-surely. Then*

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{v}{C^2} h_1\left(\frac{Ct}{v}\right)\right)$$

where $h_1(x) = (1+x) \log(1+x) - x$ for $x > 0$. Furthermore, using the inequality $h_1(x) \geq \frac{x^2}{2(1+x/3)}$ we obtain

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{t^2}{2(v + Ct/3)}\right).$$

Example. Suppose $X_i \sim \text{Bern}(p_n)$ are independent for $1 \leq i \leq n$. Then

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{2t^2}{n}\right) \quad (\text{Hoeffding})$$

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{t^2}{np_n(1-p_n) + t/3}\right). \quad (\text{Bennett})$$

Note that if $p_n \ll q$, e.g $p_n = 1/\sqrt{n}$, Hoeffding will stay the same, i.e of order $e^{-\frac{2t^2}{n}}$ (only depends on support, not variance). However, Bennet will be of the order $e^{-\frac{t^2}{\sqrt{n+t/3}}}$.