- Variance and Covariance
- Moment and Deviation
- Concentration and Tail Inequalities
- Sampling and Estimation

# Example 1: Probabilistic Packet Marking (PPM)

The Setting

- A stream of packets are sent $S = R_0 \to R_1 \to \cdots \to R_{n-1} \to D$
- Each $R_i$ can overwrite the $\text{SOURCE IP}$ field $F$ of a packet
- $D$ wants to know the set of routers on the route

The Assumption

- For each packet $D$ receives and each $i$, Prob$[F = R_i] = 1/n$ (*)

The Questions

1. How do the routers ensure (*)? Answer: Reservoir Sampling
2. How many packets must $D$ receive to know all routers?

# Coupon Collector Problem

The setting

- $n$ types of coupons
- Every cereal box has a coupon
- For each box $B$ and each coupon type $t$,

$$\text{Prob}\left[B \text{ contains coupon type } t\right] = \frac{1}{n}$$

**Coupon Collector Problem**

How many boxes of cereal must the collector purchase before he has all types of coupons?

# The Analysis

- $X$ = number of boxes he buys to have all coupon types.
- For $i \in [n]$, let $X_i$ be the additional number of cereal boxes he buys to get a new coupon type, after he had collected $i-1$ different types

$$X = X_1 + X_2 + \cdots + X_n, \quad \mathsf{E}[X] = \sum_{i=1}^{n} E[X_i]$$

- After $i-1$ types collected,

$$\text{Prob[A new box contains a new type]} = p_i = 1 - \frac{i-1}{n}$$

- Hence, $X_i$ is *geometric* with parameter $p_i$, implying

$$\mathsf{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$$

$$\mathsf{E}[X] = n \sum_{i=1}^{n} \frac{1}{n-i+1} = nH_n = n \ln n + \Theta(n)$$

# PTCF: Geometric Distribution

- A coin turns head with probability $p$, tail with $1 - p$
- $X =$ number of flips until a head shows up
- $X$ has geometric distribution with parameter $p$

$$\begin{aligned} \text{Prob}[X = n] &= (1-p)^{n-1}p \\ \text{E}[X] &= \frac{1}{p} \\ \text{Var}[X] &= \frac{1-p}{p^2} \end{aligned}$$

# Additional Questions

- We can't be sure that buying $nH_n$ cereal boxes suffices
- Want Prob$[X \geq C]$, i.e. *what's the probability that he has to buy $C$ boxes to collect all coupon types?*
- Intuitively, $X$ is far from its mean with small probability
- Want something like

$$\text{Prob}[X \geq C] \leq \text{some function of } C\text{, preferably} \ll 1$$

i.e. a (large) deviation inequality or tail inequality

## Central Theme

The more we know about $X$, the better the deviation inequality we can derive: Markov, Chebyshev, Chernoff, etc.

# PTCF: Markov's Inequality

**Theorem**

*If $X$ is a r.v. taking only non-negative values, $\mu = \mathsf{E}[X]$, then $\forall a > 0$*

$$\mathsf{Prob}[X \geq a] \leq \frac{\mu}{a}.$$

*Equivalently,*

$$\mathsf{Prob}[X \geq a\mu] \leq \frac{1}{a}.$$

If we know $\mathsf{Var}[X]$, we can do better!

- Let $X_1, \cdots, X_n$ be $n$ discrete r.v., their joint PMF is

$$p(x_1, \cdots, x_n) = \text{Prob}[X_1 = x_1 \wedge \cdots \wedge X_n = x_n].$$

- They are independent random variables iff

$$p(x_1, \cdots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n), \quad \forall x_i$$

# PTCF: (Co)Variance, Moments, Their Properties

- Variance: $\sigma^2 = \mathsf{Var}[X] := \mathsf{E}[(X - \mathsf{E}[X])^2] = \mathsf{E}[X^2] - (\mathsf{E}[X])^2$
- Standard deviation: $\sigma := \sqrt{\mathsf{Var}[X]}$
- $k$th moment: $\mathsf{E}[X^k]$
- Covariance: $\mathsf{Cov}[X, Y] := \mathsf{E}[(X - \mathsf{E}[X])(Y - \mathsf{E}[Y])]$
- For any two r.v. $X$ and $Y$,

$$\mathsf{Var}[X + Y] = \mathsf{Var}[X] + \mathsf{Var}[Y] + 2\,\mathsf{Cov}[X, Y]$$

- If $X$ and $Y$ are independent, then

$$\begin{aligned} \mathsf{E}[X \cdot Y] &= \mathsf{E}[X] \cdot \mathsf{E}[Y] \\ \mathsf{Cov}[X, Y] &= 0 \\ \mathsf{Var}[X + Y] &= \mathsf{Var}[X] + \mathsf{Var}[Y] \end{aligned}$$

- In fact, if $X_1, \dots, X_n$ are mutually independent, then

$$\mathsf{Var}\left[\sum_i X_i\right] = \sum_i \mathsf{Var}[X_i]$$

# PTCF: Chebyshev's Inequality

## Theorem (Two-sided Chebyshev's Inequality)

*If $X$ is a r.v. with mean $\mu$ and variance $\sigma^2$, then $\forall a > 0$,*

$$\text{Prob}\big[|X-\mu| \geq a\big] \leq \frac{\sigma^2}{a^2} \text{ or, equivalently } \text{Prob}\big[|X-\mu| \geq a\sigma\big] \leq \frac{1}{a^2}.$$

## Theorem (One-sided Chebyshev's Inequality)

*Let $X$ be a r.v. with $\mathsf{E}[X] = \mu$ and $\mathsf{Var}[X] = \sigma^2$, then $\forall a > 0$,*

$$\begin{aligned} \text{Prob}[X \geq \mu + a] &\leq \frac{\sigma^2}{\sigma^2 + a^2} \\ \text{Prob}[X \leq \mu - a] &\leq \frac{\sigma^2}{\sigma^2 + a^2}. \end{aligned}$$

# Back to the Additional Questions

- Markov's leads to,

$$\text{Prob}[X \geq 2nH_n] \leq \frac{1}{2}$$

- To apply Chebyshev's, we need $\text{Var}[X]$:

$$\text{Prob}[X \geq 2nH_n] \leq \text{Prob}[|X - nH_n| \geq nH_n] \leq \frac{\text{Var}[X]}{(nH_n)^2}$$

- Key observation: the $X_i$ are independent (why?)

$$\text{Var}[X] = \sum_i \text{Var}[X_i] = \sum_i \frac{1-p_i}{p_i^2} \leq \sum_i \frac{n^2}{(n-i+1)^2} = \frac{\pi^2 n^2}{6}$$

- Chebyshev's leads to

$$\text{Prob}[X \geq 2nH_n] \leq \frac{\pi^2}{6H_n^2} = \Theta\left(\frac{1}{\ln^2 n}\right)$$

# Power of Union Bound

- Chebyshev gives:

$$\text{Prob}[X \geq nH_n + cn] \leq \frac{\pi^2 n^2/6}{(cn)^2} = \Theta(1/c^2).$$

- For any fixed coupon $i$

$$\text{Prob}[i \text{ not collected after } t \text{ steps}] = \left(1 - \frac{1}{n}\right)^t \leq e^{-t/n}.$$

- Union bound gives:

$$\text{Prob}[\text{some missing coupon after } t = nH_n + cn] \leq ne^{-H_n - c} = \Theta(1/e^c).$$

# Example 2: PPM with One Bit

## The Problem

Alice wants to send to Bob a message $b_1 b_2 \cdots b_m$ of $m$ bits. She can send only **one** bit at a time, but always forgets which bits have been sent. Bob knows $m$, nothing else about the message.

## The solution

- Send bits so that the fraction of bits $1$ received is within $\epsilon$ of $p = B/2^m$, where $B = b_1 b_2 \cdots b_m$ as an integer
- Specifically, send bit $1$ with probability $p$, and $0$ with $(1 - p)$

## The question

How many bits must be sent so $B$ can be decoded with high probability?

- One way to do decoding: round the fraction of bits $1$ received to the closest multiple of of $1/2^m$
- Let $X_1, \dots, X_n$ be the bits received (independent Bernoulli trials)
- Let $X = \sum_i X_i$, then $\mu = \mathsf{E}[X] = np$. We want, say

$$\text{Prob}\left[\left|\frac{X}{n} - p\right| \leq \frac{1}{3 \cdot 2^m}\right] \geq 1 - \epsilon$$

which is equivalent to

$$\text{Prob}\left[|X - \mu| \leq \frac{n}{3 \cdot 2^m}\right] \geq 1 - \epsilon$$

This is a kind of concentration inequality.

# PTCF: The Binomial Distribution

- $n$ independent trials are performed, each with success probability $p$.
- $X =$ number of successes after $n$ trials, then

$$\text{Prob}[X = i] = \binom{n}{i} p^i (1-p)^{n-i}, \ \forall i = 0, \dots, n$$

- $X$ is called a binomial random variable with parameters $(n, p)$.

$$\begin{aligned} \mathsf{E}[X] &= np \\ \mathsf{Var}[X] &= np(1-p) \end{aligned}$$

# PTCF: Chernoff Bounds

**Theorem (Chernoff bounds are just the following idea)**

*Let $X$ be any r.v., then*

1. *For any $t > 0$*

$$\mathsf{Prob}[X \geq a] \leq \frac{\mathsf{E}[e^{tX}]}{e^{ta}}$$

   *In particular,*

$$\mathsf{Prob}[X \geq a] \leq \min_{t>0} \frac{\mathsf{E}[e^{tX}]}{e^{ta}}$$

2. *For any $t < 0$*

$$\mathsf{Prob}[X \leq a] \leq \frac{\mathsf{E}[e^{tX}]}{e^{ta}}$$

   *In particular,*

$$\mathsf{Prob}[X \geq a] \leq \min_{t<0} \frac{\mathsf{E}[e^{tX}]}{e^{ta}}$$

($\mathsf{E}^{tX}$ is called the moment generating function of $X$)

# PTCF: A Chernoff Bound for sum of Poisson Trials

**Above the mean** case.
Let $X_1, \dots, X_n$ be independent Poisson trials, $\text{Prob}[X_i = 1] = p_i$, $X = \sum_i X_i$, $\mu = \mathsf{E}[X]$. Then,

- For any $\delta > 0$,

$$\text{Prob}[X \geq (1+\delta)\mu] < \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu;$$

- For any $0 < \delta \leq 1$,

$$\text{Prob}[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/3};$$

- For any $R \geq 6\mu$,

$$\text{Prob}[X \geq R] \leq 2^{-R}.$$

# PTCF: A Chernoff Bound for sum of Poisson Trials

**Below the mean** case.

Let $X_1, \dots, X_n$ be independent Poisson trials, Prob$[X_i = 1] = p_i$, $X = \sum_i X_i$, $\mu = \mathsf{E}[X]$. Then, for any $0 < \delta < 1$:

1. $$\text{Prob}[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu};$$

2. $$\text{Prob}[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}.$$

# PTCF: A Chernoff Bound for sum of Poisson Trials

**A simple (two-sided) deviation** case.
Let $X_1, \dots, X_n$ be independent Poisson trials, Prob$[X_i = 1] = p_i$, $X = \sum_i X_i$, $\mu = \mathsf{E}[X]$. Then, for any $0 < \delta < 1$:

$$\text{Prob}[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}.$$

## Chernoff Bounds Informally

The probability that the sum of independent Poisson trials is far from the sum's mean is exponentially small.

# Back to the 1-bit PPM Problem

$$
\begin{aligned}
\text{Prob}\left[|X-\mu| > \frac{n}{3\cdot 2^m}\right] &= \text{Prob}\left[|X-\mu| > \frac{1}{3\cdot 2^m p}\mu\right] \\
&\leq \frac{2}{\exp\{\frac{n}{18\cdot 4^m p}\}}
\end{aligned}
$$

Now,

$$\frac{2}{\exp\{\frac{n}{18\cdot 4^m p}\}} \leq \epsilon$$

is equivalent to

$$n \geq 18p\ln(2/\epsilon)4^m.$$

# Example 3: A Statistical Estimation Problem

## The Problem

We want to estimate $\mu = \mathsf{E}[X]$ for some random variable $X$ (e.g., $X$ is the income in dollars of a random person in the world).

## The Question

How many samples must be take so that, given $\epsilon, \delta > 0$, the estimated value $\bar{\mu}$ satisfies

$$\mathsf{Prob}[|\overline{\mu} - \mu| \leq \epsilon\mu] \geq 1 - \delta$$

- $\delta$: confidence parameter
- $\epsilon$: error parameter
- In statistics: $[\bar{\mu}/(1+\epsilon), \bar{\mu}/(1-\epsilon)]$ is the confidence interval for $\mu$ at confidence level $1-\delta$

# Intuitively: Use "Law of Large Numbers"

- law of large numbers (there are actually 2 versions) basically says that the sample mean tends to the true mean as the number of samples tends to infinity
- We take $n$ samples $X_1, \dots, X_n$, and output

$$\bar{\mu} = \frac{1}{n}(X_1 + \cdots + X_n)$$

- But, how large must $n$ be? ("Easy" if $X$ is Bernoulli!)
- Markov is of some use, but only gives upper-tail bound
- Need a bound on the variance $\sigma^2 = \text{Var}[X]$ too, to answer the question

# Applying Chebyshev

- Let $Y = X_1 + \cdots + X_n$, then $\overline{\mu} = Y/n$ and $\mathsf{E}[Y] = n\mu$
- Since the $X_i$ are independent, $\mathsf{Var}[Y] = \sum_i \mathsf{Var}[X_i] = n\sigma^2$
- Let $r = \sigma/\mu$, Chebyshev inequality gives

$$\begin{aligned} \mathsf{Prob}[|\overline{\mu} - \mu| > \epsilon\mu] &= \mathsf{Prob}\left[|Y - \mathsf{E}[Y]| > \epsilon \mathsf{E}[Y]\right] \\ &< \frac{\mathsf{Var}[Y]}{(\epsilon \mathsf{E}[Y])^2} = \frac{n\sigma^2}{\epsilon^2 n^2 \mu^2} = \frac{r^2}{n\epsilon^2}. \end{aligned}$$

- Consequently, $n = \frac{r^2}{\delta\epsilon^2}$ is sufficient!
- We can do better!

# Finally, the Median Trick!

- If confident parameter is $1/4$, we only need $\Theta(r^2/\epsilon^2)$ samples; the estimate is a little "weak"
- Suppose we have $w$ weak estimates $\mu_1, \dots, \mu_w$
- Output $\bar{\mu}$: the **median** of these weak estimates!
- Let $I_j$ indicates the event $|\mu_j - \mu| \leq \epsilon\mu$, and $I = \sum_{j=1}^{w} I_j$
- By Chernoff's bound,

$$\begin{aligned}
\text{Prob}[|\overline{\mu} - \mu| > \epsilon\mu] &\leq \text{Prob}\left[Y \leq w/2\right] \\
&\leq \text{Prob}\left[Y \leq (2/3)\mathsf{E}[Y]\right] \\
&= \text{Prob}\left[Y \leq (1 - 1/3)\mathsf{E}[Y]\right] \\
&\leq \frac{1}{e^{\mathsf{E}[Y]/18}} \leq \frac{1}{e^{w/24}} \leq \delta
\end{aligned}$$

  whenever $w \geq 24 \ln(1/\delta)$.
- Thus, the total number of samples needed is $n = O(r^2 \ln(1/\delta)/\epsilon^2)$.