CMPT 210: Probability and Computation

Lecture 17

Sharan Vaswani

July 12, 2022

Collect Assignment 2

Recap

Standard Deviation: For r.v. X, the standard deviation of X is defined as $\sigma_X := \sqrt{\mathrm{Var}[X]} = \sqrt{\mathbb{E}[X^2] - (\mathbb{E}[X])^2}$.

For constants a, b and r.v. R, $\mathrm{Var}[aR + b] = a^2 \mathrm{Var}[R]$.

Pairwise Independence: Random variables $R_1, R_2, R_3, \ldots R_n$ are pairwise independent if for any pair R_i and R_j, for $x \in \mathrm{Range}(R_i)$ and $y \in \mathrm{Range}(R_j)$, $\Pr[(R_i = x) \cap (R_j = y)] = \Pr[R_i = x] \Pr[R_j = y]$.

Linearity of variance for pairwise independent r.v's: If $R_1, \ldots, R_n$ are pairwise independent, $\mathrm{Var}[R_1 + R_2 + \ldots R_n] = \sum_{i=1}^n \mathrm{Var}[R_i]$.

Matching Birthdays

Q: In a class of n students, what is the probability that two students share the same birthday? Assume that (i) each student is equally likely to be born on any day of the year, (ii) no leap years and (iii) student birthdays are independent of each other.

For $d := 365$,

$$\Pr[\text{two students share the same birthday}] = 1 - \frac{d \times (d-1) \times (d-2) \times \ldots (d-(n-1))}{d^n}$$

Q: On average, how many pairs of students have matching birthdays?

Define M to be the number of pairs of students with matching birthdays. For a fixed ordering of the students, let $X_{i,j}$ be the indicator r.v. corresponding to the event $E_{i,j}$ that the birthdays of students i and j match. Hence,

$$M = \sum_{i,j|1 \leq i < j \leq n} X_{i,j} \implies \mathbb{E}[M] = \mathbb{E}[\sum_{i,j|1 \leq i < j \leq n} X_{i,j}] = \sum_{i,j|1 \leq i < j \leq n} \mathbb{E}[X_{i,j}] = \sum_{i,j|1 \leq i < j \leq n} \Pr[E_{i,j}]$$

(Linearity of expectation)

Matching Birthdays

For a pair of students i, j, let B_i be the r.v. equal to the day of student i's birthday. Range(B_i) $= \{1,2,\ldots, 365\}$ and for all $k \in [365]$, $\Pr[B_i = k] = 1/d$.

$$E_{i,j} = (B_i = 1 \cap B_j = 1) \cup (B_i = 2 \cap B_j = 2) \cup \ldots$$

$$\implies \Pr[E_{i,j}] = \sum_{k=1}^{d} \Pr[B_i = k \cap B_j = k] = \sum_{k=1}^{d} \Pr[B_i = k] \Pr[B_j = k] = \sum_{k=1}^{d} \frac{1}{d^2} = \frac{1}{d}$$

$$\implies \mathbb{E}[M] = \sum_{i,j|1 \le i < j \le n} \Pr[E_{i,j}] = \frac{1}{d} \sum_{i,j|1 \le i < j \le n} (1) = \frac{1}{d}[(n-1) + (n-2) + \ldots + 1] = \frac{n(n-1)}{2d}$$

Hence, in our class of 48 students, on average, there are $\frac{(24)(47)}{365} = 3.09$ students with matching birthdays.

Matching Birthdays

Q: Are the $X_{i,j}$ mutually independent?

No, because if $X_{i,j} = 1$ and $X_{j,k} = 1$, then,
$\Pr[X_{i,k} = 1 | X_{j,k} = 1 \cap X_{i,j} = 1] = 1 \neq \frac{1}{d} = \Pr[X_{i,k} = 1]$.

Q: Are the $X_{i,j}$ pairwise independent?

Yes, because for all i, j and i', j' (where $i \neq i'$), $\Pr[X_{i,j} = 1 | X_{i',j'} = 1] = \Pr[X_{i,j} = 1]$ because if students i' and j' have matching birthdays, it does not tell us anything about whether i and j have matching birthdays.

Matching Birthdays

Q: If M is the r.v. equal to the number of pairs of students with matching birthdays, calculate $\text{Var}[M]$.

$$\text{Var}[M] = \text{Var}[\sum_{i,j|1 \leq i < j \leq n} X_{i,j}]$$

Since $X_{i,j}$ are pairwise independent, the variance of the sum is equal to the sum of the variance.

$$\implies \text{Var}[M] = \sum_{i,j|1 \leq i < j \leq n} \text{Var}[X_{i,j}] = \sum_{i,j|1 \leq i < j \leq n} \frac{1}{d}\left(1 - \frac{1}{d}\right) = \frac{1}{d}\left(1 - \frac{1}{d}\right)\frac{n(n-1)}{2}$$

(Since $X_{i,j}$ is an indicator (Bernoulli) r.v.)

Hence, in our class of 48 students, the standard deviation for the matching birthdays is equal to $\sqrt{\frac{(24)(47)}{365} \frac{364}{365}} \approx 1.75$.

Questions?

Covariance

For two random variables R and S, the covariance between R and S is defined as:

$$\text{Cov}[R,S] = \mathbb{E}[(R - \mathbb{E}[R])\,(S - \mathbb{E}[S])] = \mathbb{E}[RS] - \mathbb{E}[R]\,\mathbb{E}[S]$$

$$\begin{aligned}
\text{Cov}[R,S] &= \mathbb{E}[(R - \mathbb{E}[R])\,(S - \mathbb{E}[S])] \\
&= \mathbb{E}\left[RS - R\,\mathbb{E}[S] - S\,\mathbb{E}[R] + \mathbb{E}[R]\,\mathbb{E}[S]\right] \\
&= \mathbb{E}[RS] - \mathbb{E}[R\,\mathbb{E}[S]] - \mathbb{E}[S\,\mathbb{E}[R]] + \mathbb{E}[R]\,\mathbb{E}[S] \\
\implies \text{Cov}[R,S] &= \mathbb{E}[RS] - \mathbb{E}[R]\,\mathbb{E}[S] - \mathbb{E}[S]\,\mathbb{E}[R] + \mathbb{E}[R]\,\mathbb{E}[S] = \mathbb{E}[RS] - \mathbb{E}[R]\,\mathbb{E}[S]
\end{aligned}$$

Covariance generalizes the notion of variance to multiple random variables.

$$\text{Cov}[R,R] = \mathbb{E}[R\,R] - \mathbb{E}[R]\,\mathbb{E}[R] = \text{Var}[R]$$

If R and S are independent r.v's, $\mathbb{E}[RS] = \mathbb{E}[R]\,\mathbb{E}[S]$ and $\text{Cov}[R,S] = 0$.

The covariance between two r.v's is symmetric i.e. $\text{Cov}[R,S] = \text{Cov}[S,R]$.

Covariance

For two arbitrary (not necessarily independent) r.v's, R and S,

$$\mathrm{Var}[R+S] = \mathrm{Var}[R] + \mathrm{Var}[S] + 2\,\mathrm{Cov}[R,S]$$

Recall that,

$$\mathrm{Var}[R+S] = \mathrm{Var}[R] + \mathrm{Var}[S] + 2(\mathbb{E}[RS] - \mathbb{E}[R]\mathbb{E}[S]) = \mathrm{Var}[R] + \mathrm{Var}[S] + 2\,\mathrm{Cov}[R,S]$$

If R and S are independent, $\mathrm{Cov}[R,S] = 0$ and we recover the formula for the sum of independent variables.

For $R = S$, $\mathrm{Var}[R+R] = \mathrm{Var}[R] + \mathrm{Var}[R] + 2\mathrm{Cov}[R,R] = \mathrm{Var}[R] + \mathrm{Var}[R] + 2\mathrm{Var}[R] = 4\mathrm{Var}[R]$ which is consistent with our previous formula that $\mathrm{Var}[2R] = 2^2\mathrm{Var}[R]$.

Generalization to multiple random variables $R_1, R_2, \ldots R_n$:

$$\mathrm{Var}\left[\sum_{i=1}^{n} R_i\right] = \sum_{i=1}^{n} \mathrm{Var}[R_i] + 2 \sum_{1 \le i < j \le n} \mathrm{Cov}[R_i, R_j]$$

Covariance - Example

Q: If X and Y are indicator r.v's for events A and B respectively, calculate the covariance between X and Y

Note that $X = \mathcal{I}_A$ and $Y = \mathcal{I}_B$ and $XY = \mathcal{I}_{A \cap B}$.

$$\mathbb{E}[X] = \Pr[A]\,; \mathbb{E}[Y] = \Pr[B]$$

$$\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = \Pr[A \cap B] - \Pr[A]\Pr[B]$$

If $\mathrm{Cov}[X, Y] > 0 \implies \Pr[A \cap B] > \Pr[A]\Pr[B]$. Hence,

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]} > \frac{\Pr[A]\Pr[B]}{\Pr[B]} = \Pr[A]$$

If $Cov[X, Y] > 0$, it implies that $\Pr[A|B] > \Pr[A]$ and hence, the probability that event A happens increases if B is going to happen/has happened. Similarly, if $Cov[X, Y] < 0$,

$$\Pr[A|B] < \Pr[A]$$

In this case, if B happens, then the probability of event A decreases.

Correlation

The correlation between two r.v's R_1 and R_2 is defined as:

$$\mathrm{Corr}[R_1, R_2] = \frac{\mathrm{Cov}[R_1, R_2]}{\sqrt{\mathrm{Var}[R_1]\,\mathrm{Var}[R_2]}}$$

$\mathrm{Corr}[R_1, R_2] \in [-1, 1]$ and indicates the strength of the relationship between R_1 and R_2.

If $\mathrm{Corr}[R_1, R_2] > 0$, then R_1 and R_2 are said to be positively correlated, else if $\mathrm{Corr}[R_1, R_2] < 0$, the r.v's are negatively correlated.

If $R_1 = R_2 = R$, then, $\mathrm{Corr}[R, R] = \frac{\mathrm{Cov}[R,R]}{\sqrt{\mathrm{Var}[R]\,\mathrm{Var}[R]}} = \frac{\mathrm{Var}[R]}{\mathrm{Var}[R]} = 1$.

If R_1 and R_2 are independent, $\mathrm{Cov}[R, R] = 0$ and $\mathrm{Corr}[R, R] = 0$.

If $R_1 = -R_2 = R$, then,

$$\begin{aligned}\mathrm{Corr}[R, -R] &= \frac{\mathrm{Cov}[R, -R]}{\sqrt{\mathrm{Var}[R]\,\mathrm{Var}[-R]}} = \frac{\mathrm{Cov}[R, -R]}{\sqrt{\mathrm{Var}[R]\,(-1)^2\mathrm{Var}[R]}} = \frac{\mathrm{Cov}[R, -R]}{\mathrm{Var}[R]} \\ &= \frac{\mathbb{E}[-R^2] - \mathbb{E}[R]\,\mathbb{E}[-R]}{\mathrm{Var}[R]} = \frac{-\mathbb{E}[R^2] + \mathbb{E}[R]\,\mathbb{E}[R]}{\mathrm{Var}[R]} = \frac{-\mathrm{Var}[R]}{\mathrm{Var}[R]} = -1\end{aligned}$$

Questions?

Tail inequalities

Variance gives us one way to measure how "spread" the distribution is – weighted average of the deviation of the random variable from its mean.

Tail inequalities will help give a more precise characterization of the deviation from the mean.

Typically, tail inequalities give bounds on the probability that the r.v. takes a value much different from its mean.

Consider a r.v. X that can take on only non-negative values and $\mathbb{E}[X] = 100$. This immediately implies that $\Pr[X \geq 300] \leq \frac{1}{3}$. Since if $\Pr[X \geq 300] > \frac{1}{3}$,
$\mathbb{E}[X] > (300)\,\frac{1}{3} +$ other positive quantities > 100 which is a contradiction.

Similarly, $\Pr[X \geq 150] \leq \frac{2}{3}$. Hence, we can bound the probability that X takes on values too far away from its mean (equal to 100) by using the knowledge of $\mathbb{E}[X]$ and the fact that X can take only non-negative values.

Markov's Theorem

Markov's theorem formalizes the intuition on the previous slide, and can be stated as follows.

Markov's Theorem: If X is a non-negative random variable, then for all $x > 0$,

$$\Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}.$$

Define $\mathcal{I}_x$ to be the indicator r.v. for the event $[X \geq x]$. Then for all values of X, $x\,\mathcal{I}_x \leq X$. Taking expectations,

$$\mathbb{E}[x\,\mathcal{I}_x] \leq \mathbb{E}[X] \implies x\,\mathbb{E}[\mathcal{I}_x] \leq \mathbb{E}[X] \implies x\Pr[X \geq x] \leq \mathbb{E}[X] \implies \Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}.$$

Since the above theorem holds for all $x > 0$, let's set $x = c\mathbb{E}[X]$ for $c \geq 1$. Hence,

$$\Pr[X \geq c\mathbb{E}[X]] \leq \frac{1}{c}$$

Hence, the probability that X is "far" from the mean in terms of the multiplicative factor c is upper-bounded by $\frac{1}{c}$.

Markov's Theorem – Example

Q: Suppose there is a dinner party where n people check in their coats. The coats are mixed up during dinner, so that afterward each person receives a random coat. In particular, each person gets their own coat with probability $\frac{1}{n}$.

Recall that if G is the r.v. corresponding to the number of people that receive their own coat, then we used the linearity of expectation to derive that $\mathbb{E}[G] = 1$. Using Markov's Theorem,

$$\Pr[G \geq x] \leq \frac{\mathbb{E}[G]}{x} = \frac{1}{x}.$$

Hence, we can bound the probability that x people receive their own coat. For example, there is no better than 20% chance that 5 people get their own coat.

Markov's Theorem – Example

Q: Suppose n people are eating different appetizers arranged on a circular, rotating banquet tray. Someone then spins the tray so that each person receives a random appetizer. What is the probability that everyone gets the same appetizer as before?

There are n possible orientations for the tray, and hence the probability that everyone gets the same appetizer is $\frac{1}{n}$. Let us solve this using Markov's Theorem.

If R is the r.v. corresponding to the number of people who get the same appetizer, $\mathbb{E}[R] = (n)\left(\frac{1}{n}\right) + (0)\left(1 - \frac{1}{n}\right) = 1$. Using Markov's Theorem with $x = n$,

$$\Pr[R \geq n] \leq \frac{\mathbb{E}[R]}{n} = \frac{1}{n}.$$

Hence, Markov's inequality is "tight" for this example, and exactly gives the probability that $\Pr[R = n]$.

Markov's Theorem – Example

Q: If X is a non-negative r.v. such that $\mathbb{E}[X] = 150$, compute the probability that X is at least 200.

If we know that X can not take values less than 100, we can use Markov's Theorem to get a tighter bound.

Define $Y := X - 100$. $\mathbb{E}[Y] = \mathbb{E}[X] - 100 = 50$ and Y is non-negative.

$$\Pr[X \geq 200] = \Pr[Y + 100 \geq 200] = \Pr[Y \geq 100] \leq \frac{\mathbb{E}[Y]}{100} = \frac{50}{100} = \frac{1}{2}$$

Hence, if we have additional information (in the form of a lower-bound that a r.v. can not be smaller than some constant $b > 0$), we can use Markov's Theorem on the shifted r.v. (Y in our example) and obtain a tighter bound on the probability of deviation. We will make this precise in Assignment 4!

Chebyshev's Theorem

General Idea: Use Markov's Theorem with some cleverly chosen function of X. Formally, for some function f such that $Y := f(X)$ is non-negative. Using Markov's Theorem for Y,

$$\Pr[f(X) \geq x] \leq \frac{\mathbb{E}[f(X)]}{x}$$

Choosing $f(X) = |X - \mathbb{E}[X]|^2$ and $x = y^2$ implies that $f(X)$ is non-negative and $x > 0$. Using Markov's Theorem,

$$\Pr[|X - \mathbb{E}[X]|^2 \geq y^2] \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{y^2}$$

Note that $\Pr[|X - \mathbb{E}[X]|^2 \geq y^2] = \Pr[|X - \mathbb{E}[X]| \geq y]$, and hence,

$$\Pr[|X - \mathbb{E}[X]| \geq y] \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{y^2} = \frac{\mathrm{Var}[X]}{y^2}$$

Chebyshev's Theorem: For a r.v. X and a constant $x > 0$,

$$\Pr[|X - \mathbb{E}[X]| \geq x] \leq \frac{\mathrm{Var}[X]}{x^2}.$$

Chebyshev's Theorem

Chebyshev's Theorem bounds the probability that the random variable X is "far" away from the mean $\mathbb{E}[X]$ by an additive factor of x.

If we set $x = c\sigma_X$ where σ_X is the standard deviation of X, then by Chebyshev's Theorem,

$$\Pr[|X - \mathbb{E}[X]| \geq c\sigma_X] \leq \frac{\mathrm{Var}[X]}{c^2\sigma_X^2} = \frac{1}{c^2}$$

Hence,

$$\Pr[\mathbb{E}[X] - c\sigma_X < X < \mathbb{E}[X] + c\sigma_X] = 1 - \Pr[|X - \mathbb{E}[X]| \geq c\sigma_X] \geq 1 - \frac{1}{c^2}.$$

Hence, Chebyshev's Theorem can be used to bound the probability that X is "concentrated" near its mean.

Chebyshev's Theorem - Example

Q: If X is a non-negative r.v. such that $\mathbb{E}[X] = 100$ and $\sigma_X = 15$, compute the probability that X is at least 300.

If we use Markov's Theorem, $\Pr[X \geq 300] \leq \frac{\mathbb{E}[X]}{300} = \frac{1}{3}$.

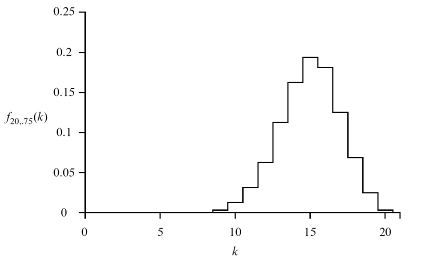
Note that $\Pr[|X - 100| \geq 200] = \Pr[X \leq -100 \cup X \geq 300] = \Pr[X \geq 300]$. Using Chebyshev's Theorem,

$$\Pr[|X - 100| \geq 200] \leq \frac{\text{Var}[X]}{(200)^2} = \frac{15^2}{200^2} \approx \frac{1}{178}.$$

Hence, by exploiting the knowledge of the variance and using Chebyshev's inequality, we can obtain a tighter bound.

Chebyshev's Theorem - Example

Q: Consider an r.v. $X \sim \text{Bin}(20, 0.75)$. Plot the PDF_X, compute its mean and standard deviation and bound $\Pr[10 < X < 30]$.



$\text{Range}(X) = \{0, 1, \ldots, 20\}$ and for $k \in \text{Range}(X)$, $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

$\mathbb{E}[X] = np = (20)(0.75) = 15$

$\text{Var}[X] = np(1-p) = 20(0.75)(0.25) = 3.75$ and hence $\sigma_X = \sqrt{3.75} \approx 1.94$.

$\Pr[10 < X < 30] = 1 - \Pr[X \leq 10 \cup X \geq 30] = 1 - \Pr[|X - 20| \geq 10] = 1 - \frac{\text{Var}[X]}{(10)^2} \geq 1 - \frac{3.75}{100} = 1 - 0.0375 = 0.9625.$

Hence, the "probability mass" of X is "concentrated" around its mean.

Chebyshev's Theorem - Example

Q: In a class of n students, assume that (i) each student is equally likely to be born on any day of the year, (ii) no leap years and (iii) student birthdays are independent of each other, if M is the r.v. equal to the number of pairs of students with matching birthdays, calculate $\Pr[|M - \mathbb{E}[M]| > x]$ for $n = 48$.

Recall that for $n = 48$, $\mathbb{E}[M] \approx 3.09$ and $\mathrm{Var}[M] \approx 3.08$. Hence, by Chebyshev's Theorem,

$$\Pr[|M - 3.09| > x] \leq \frac{3.08}{x^2}.$$

Hence, for $x = 3$, $\Pr[|M - 3.09| > 3] \leq \frac{3.08}{9} \approx 0.34$. Hence, there is 34% chance that the number of matched birthdays is greater than 6.09 and smaller than 0.09.

Questions?

Voter Poll

Q: Suppose there is an election between two candidates A and B, and we are hired by candidate A's election campaign to estimate the chances of A winning the election. In particular, we want to estimate p, the fraction of voters favoring A before the election. We conduct a voter poll – selecting (typically calling) people uniformly at random (with replacement so that we can choose a person twice) and try to estimate p. What is the number of people we should poll to estimate p reasonably accurately and with reasonably high probability?

Let us define X_i to be the indicator r.v. which is equal to 1 if person i that we called favors candidate A. The X_i r.v's are mutually independent since the people we poll are chosen randomly, and we assume that they are identically distributed meaning that $X_i = 1$ with probability p.

Suppose we poll n people and define $S_n := \sum_{i=1}^{n} X_i$ as the r.v. equal to the total number of people who prefer candidate A (amongst the people we polled). $\frac{S_n}{n}$ is the fraction of polled voters who favor candidate A and is the *statistical estimate* of p.

Q: What is the distribution of S_n?

Voter Poll

Hence, we want to find for what n is our estimate for p accurate up to an error $\epsilon > 0$ and with probability $1 - \delta$ (for $\delta \in (0, 1)$). Formally, for what n is,

$$\Pr\left[\left|\frac{S_n}{n} - p\right| < \epsilon\right] \geq 1 - \delta$$

Since $S_n \sim \text{Bin}(n, p)$, $\mathbb{E}[S_n] = np$ and hence, $\mathbb{E}\left[\frac{S_n}{n}\right] = p$, meaning that our estimate is *unbiased* – in expectation, the estimate is equal to p. Hence, the above statement is equivalent to,

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| < \epsilon\right] \geq 1 - \delta$$

Hence, we can use Chebyshev's Theorem for the r.v. $\frac{S_n}{n}$ with $x = \epsilon$ to bound the LHS

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| < \epsilon\right] = 1 - \Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| \geq \epsilon\right] \geq 1 - \frac{\text{Var}[S_n/n]}{\epsilon^2}.$$

Hence, the problem now is to find n such that,

$$1 - \frac{\text{Var}[S_n/n]}{\epsilon^2} \geq 1 - \delta \implies \frac{\text{Var}[S_n/n]}{\epsilon^2} < \delta$$

Voter Poll

Let us calculate the $\mathrm{Var}[S_n/n]$.

$$\begin{aligned}\mathrm{Var}[S_n/n] &= \frac{1}{n^2}\mathrm{Var}[S_n] && \text{(Using the property of variance)}\\ &= \frac{1}{n^2}\, n\, p\,(1-p) = \frac{p\,(1-p)}{n} && \text{(Using the variance of the Binomial distribution)}\end{aligned}$$

Hence, we want to find n s.t.

$$\frac{p\,(1-p)}{n\epsilon^2} < \delta \implies n \geq \frac{p(1-p)}{\epsilon^2\,\delta}$$

But we do not know p! If $n \geq \max_p \frac{p(1-p)}{\epsilon^{\mathbf{2}}\,\delta}$, then for any p, $n \geq \frac{p(1-p)}{\epsilon^{\mathbf{2}}\,\delta}$. So the problem is to compute $\max_p \frac{p(1-p)}{\epsilon^{\mathbf{2}}\,\delta}$. This is a concave function and is maximized at $p = 1/2$.

Hence, if $n \geq \frac{1}{4\epsilon^{\mathbf{2}}\delta}$, then $\Pr\left[\left|\frac{S_n}{n} - p\right| < \epsilon\right] \geq 1 - \delta$ meaning that we have estimated p upto an error ϵ and this bound is true with high probability equal to $1 - \delta$.

For example, if $\epsilon = 0.01$ and $\delta = 0.01$ meaning that we want the bound to hold 99% of the time, then, we require $n \geq 250000$.

Pairwise Independent Sampling

Let $G_1, G_2, \ldots, G_n$ be pairwise independent random variables with the same mean μ and standard deviation σ. Define $S_n := \sum_{i=1}^{n} G_i$, then,

$$\Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] \leq \frac{1}{n}\left(\frac{\sigma}{\epsilon}\right)^2.$$

Let us compute $\mathbb{E}[S_n/n]$ and $\mathrm{Var}[S_n/n]$.

$$\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{i=1}^{n} G_i\right] = \sum_{i=1}^{n} \mathbb{E}[G_i] = n\mu \implies \mathbb{E}[S_n/n] = \frac{1}{n}\mathbb{E}[S_n] = \mu$$

(Using linearity of expectation)

$$\mathrm{Var}[S_n] = \mathrm{Var}\left[\sum_{i=1}^{n} G_i\right] = \sum_{i=1}^{n} \mathrm{Var}[G_i] = n\sigma^2$$

(Using linearity of variance for pairwise independent r.v's)

$$\implies \mathrm{Var}[S_n/n] = \frac{1}{n^2}\mathrm{Var}[S_n] = \frac{\sigma^2}{n}$$

Pairwise Independent Sampling

Using Chebyshev's Theorem,

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| \geq \epsilon\right] = \Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] \leq \frac{\mathrm{Var}[S_n/n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Hence, for arbitrary pairwise independent r.v's, if n increases, the probability of deviation from the mean μ decreases.

Weak Law of Large Numbers: Let $G_1, G_2, \ldots, G_n$ be pairwise independent variables with the same mean μ and (finite) standard deviation σ. Define $T_n := \frac{\sum_{i=1}^{n} G_i}{n}$, then for every $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr[|T_n - \mu| \leq \epsilon] = 1.$$

Follows from the theorem on pairwise independent sampling since $\lim_{n\to\infty} \Pr[|T_n - \mu| \leq \epsilon] = \lim_{n\to\infty} \left[1 - \frac{\sigma^2}{n\epsilon^2}\right] = 1.$

Questions?

Sums of Random Variables

If we know that the r.v X is (i) non-negative and (ii) $\mathbb{E}[X]$, we can use Markov's Theorem to bound the probability of deviation from the mean.

If we know both (i) $\mathbb{E}[X]$ and (ii) $\mathrm{Var}[X]$, we can use Chebyshev's Theorem to bound the probability of deviation.

In many cases (the voter poll example), we know the distribution of the r.v. (for voter poll, $S_n \sim \mathrm{Bin}(n, p)$) and can obtain tighter bounds on the deviation from the mean.

Chernoff Bound: Let $T_1, T_2, \ldots, T_n$ be mutually independent r.v's such that $0 \leq T_i \leq 1$ for all i. If $T := \sum_{i=1}^{n} T_i$, then, for all $c \geq 1$,

$$\Pr[T \geq c\mathbb{E}[T]] \leq \exp(-\beta(c)\,\mathbb{E}[T])$$

where, $\beta(c) := c\ln(c) - c + 1$.

If $T_i \sim \mathrm{Ber}(p)$ and are mutually independent, then $T_i \in \{0, 1\}$ and we can use the Chernoff bound to bound the deviation from the mean for $T \sim \mathrm{Bin}(n, p)$. In general, if $T_i \in [0, 1]$, the Chernoff Bound can be used even if the T_i's have different distributions!

Chernoff Bound – Binomial Distribution

Q: Bound the probability that the number of heads that come up in 1000 independent tosses of a fair coin exceeds the expectation by 20% or more.

Let T_i be the r.v. for the event that coin i comes up heads, and let T denote the total number of heads. Hence, $T = \sum_{i=1}^{1000} T_i$. For all i, $T_i \in \{0, 1\}$ and are mutually independent r.v's. Hence, we can use the Chernoff Bound.

We want to compute the probability that the number of heads is larger than the expectation by 20% meaning that $c = 1.2$ for the Chernoff Bound. Computing $\beta(c) = c\ln(c) - c + 1 \approx 0.0187$. Since the coin is fair, $\mathbb{E}[T] = 1000\,\frac{1}{2} = 500$. Plugging into the Chernoff Bound,

$$\Pr[T \geq c\mathbb{E}[T]] \leq \exp(-\beta(c)\mathbb{E}[T]) \implies \Pr[T \geq 1.2\mathbb{E}[T]] \leq \exp(-(0.0187)\,(500)) \approx 0.0000834.$$

Comparing this to using Chebyshev's inequality,

$$\begin{aligned}\Pr[T \geq c\mathbb{E}[T]] = \Pr[T - \mathbb{E}[T] \geq (c-1)\mathbb{E}[T]] &\leq \Pr[|T - \mathbb{E}[T]| \geq (c-1)\,\mathbb{E}[T]] \\ &\leq \frac{\mathrm{Var}[T]}{(c-1)^2\,(\mathbb{E}[T])^2} = \frac{1000\,\frac{1}{4}}{(1.2-1)^2(500^2)} = \frac{250}{0.2^2\,500^2} = \frac{250}{10000} = 0.025.\end{aligned}$$

Chernoff Bound – Lottery Game

Q: Pick-4 is a lottery game in which you pay \$1 to pick a 4-digit number between 0000 and 9999. If your number comes up in a random drawing, then you win \$5,000. Your chance of winning is 1 in 10000. If 10 million people play, then the expected number of winners is 1000. When there are 1000 winners, the lottery keeps \$5 million of the \$10 million paid for tickets. The lottery operator's nightmare is that the number of winners is much greater – especially at the point where more than 2000 win and the lottery must pay out more than it received. What is the probability that will happen?

Let T_i be an indicator for the event that player i wins. Then $T := \sum_{i=1}^{n} T_i$ is the total number of winners. If we assume that the players' picks and the winning number are random, independent and uniform, then the indicators T_i are independent, as required by the Chernoff bound.

We wish to compute $\Pr[T \geq 2000] = \Pr[T \geq 2\mathbb{E}[T]]$. Hence $c = 2$ and $\beta(c) \approx 0.386$. By the Chernoff bound,

$$\Pr[T \geq 2\mathbb{E}[T]] \leq \exp(-\beta(c)\mathbb{E}[T]) = \exp(-(0.386)\,1000) < \exp(-386) \approx 10^{-168}$$