CMPT 210: Probability and Computing

Lecture 19

Sharan Vaswani

March 17, 2023

Recap

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

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

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 $x \in \text{Range}(R_i)$ and $y \in \text{Range}(R_j)$, $\Pr[(R_i = x) \cap (R_j = y)] = \Pr[R_i = x] \Pr[R_j = y]$.

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

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 $d := 365$ (since no leap years),

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

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,. . . , d}. For all $k \in [d]$, $\Pr[B_i = k] = 1/d$ (each student is equally likely to be born on any day of the year).

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

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

(student birthdays are independent of each other)

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

Hence, in our class of 42 students, on average, there are $\frac{(21)(41)}{365} = 2.35$ students with matching birthdays.

Matching Birthdays

Q: Are the $X_{i,j}$ r.v's 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 random variable 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. and $\Pr[X_{i,j} = 1] = \frac{1}{d}$)

Hence, in our class of 42 students, the standard deviation for the matching birthdays is equal to $\sqrt{\frac{(21)\,(41)}{365}\,\frac{364}{365}} \approx 1.53$.

Questions?

Covariance

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

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

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

Covariance generalizes the notion of variance to multiple random variables.

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

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

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

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 from Lecture 17, Slide 7, where we showed 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 $R_1, R_2, \ldots R_n$ (Recall from Lecture 17, Slide 8):

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

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

We know that $\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$. Note that $X = \mathcal{I}_A$ and $Y = \mathcal{I}_B$. We can conclude that $XY = \mathcal{I}_{A\cap B}$ since $XY = 1$ iff both events A and B happen.

$$\implies \mathbb{E}[X] = \Pr[A]\ ; \mathbb{E}[Y] = \Pr[B]; \mathbb{E}[XY] = \Pr[A \cap B]$$

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

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

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.

Tail inequalities bound the probability that the r.v. takes a value much different from its mean.

Example: Consider a r.v. X that can take on only non-negative values and $\mathbb{E}[X] = 99.99$. Show that $\Pr[X \geq 300] \leq \frac{1}{3}$.

$$\begin{aligned} \textit{Proof}: \mathbb{E}[X] &= \sum_{x \in \text{Range}(X)} x \Pr[X = x] = \sum_{x | x \geq 300} x \Pr[X = x] + \sum_{x | 0 \leq x < 300} x \Pr[X = x] \\ &\geq \sum_{x | x \geq 300} (300) \Pr[X = x] + \sum_{x | 0 \leq x < 300} x \Pr[X = x] \\ &= (300) \Pr[X \geq 300] + \sum_{x | 0 \leq x < 300} x \Pr[X = x] \end{aligned}$$

If $\Pr[X \geq 300] > \frac{1}{3}$, then, $\mathbb{E}[X] > (300)\,\frac{1}{3} + \sum_{x|0 \leq x < 300} x \Pr[X = x] > 100$ (since the second term is always non-negative). Hence, if $\Pr[X \geq 300] > \frac{1}{3}$, $\mathbb{E}[X] > 100$ which is a contradiction since $\mathbb{E}[X] = 99.99$.

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 $x > 0$,

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

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

Q: If someone tell us that X can not take values less than 100 and $\mathbb{E}[X] = 150$, compute the probability that X is at least 200.

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.

Chebyshev's Theorem

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

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

Proof: 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

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, $\Pr[X \geq 300] \leq \frac{\mathbb{E}[X]}{300} = \frac{1}{3}$.

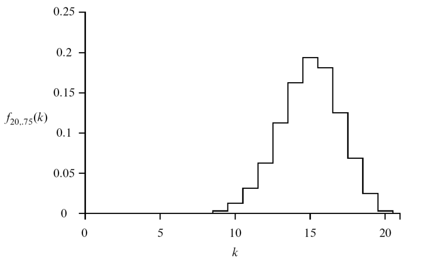
Note that $\Pr[|X - 100| \geq 200] = \Pr[X \leq -100 \cup X \geq 300] = \Pr[X \geq 300]$. Using Chebyshev's Theorem,

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

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 $\Pr[10 < X < 20]$.



Range$(X) = \{0, 1, \ldots, 20\}$ and for $k \in$ 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
$\sigma_X = \sqrt{3.75} \approx 1.94$.

$$\begin{aligned}
\Pr[10 < X < 20] &= 1 - \Pr[X \leq 10 \cup X \geq 20] \\
&= 1 - \Pr[|X - 15| \geq 5] \\
&= 1 - \Pr[|X - \mathbb{E}[X]| \geq 5] \\
&\geq 1 - \frac{\text{Var}[X]}{(5)^2} = 1 - \frac{3.75}{25} = 0.85.
\end{aligned}$$

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?