CMPT 210: Probability and Computing

Lecture 19

Sharan Vaswani

March 21, 2024

Recap

Variance: Standard way to measure the deviation from the mean. For r.v. X, $\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in \text{Range}(X)} (x - \mu)^2 \Pr[X = x]$, where $\mu := \mathbb{E}[X]$.

Alternate Definition: $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

If $X \sim \text{Ber}(p)$, $\text{Var}[X] = p(1 - p)$.

If $X \sim \text{Uniform}(\{v_1, v_2, \ldots v_n\})$, $\text{Var}[X] = \frac{[v_1^2 + v_2^2 + \ldots v_n^2]}{n} - \left(\frac{[v_1 + v_2 + \ldots v_n]}{n}\right)^2$.

Variance - Examples

Q: If $R \sim \text{Geo}(p)$, calculate $\text{Var}[R]$.

$$\text{Var}[R] = \mathbb{E}[R^2] - (\mathbb{E}[R])^2 = \mathbb{E}[R^2] - \frac{1}{p^2}$$

Recall that for a coin s.t. $\Pr[\text{heads}] = p$, R is the r.v. equal to the number of coin tosses we need to get the first heads. Let A be the event that we get a heads in the first toss. Using the law of total expectation,

$$\mathbb{E}[R^2] = \mathbb{E}[R^2|A]\Pr[A] + \mathbb{E}[R^2|A^c]\Pr[A^c]$$

$\mathbb{E}[R^2|A] = 1$ ($R^2 = 1$ if we get a heads in the first coin toss) and $\Pr[A] = p$. Hence,

$$\mathbb{E}[R^2] = (1)\,(p) + \mathbb{E}[R^2|A^c]\,(1-p) \quad ; \quad \mathbb{E}[R^2|A^c] = \sum_{k=1} k^2 \Pr[R = k|A^c]$$

Note that $\Pr[R = k|A^c] = \Pr[R = k|\text{ first toss is a tails}] = (1-p)^{k-2}\,p = \Pr[R = k-1]$

$$\implies \mathbb{E}[R^2|A^c] = \sum_{k=1} k^2 \Pr[R = k-1] = \sum_{t=0} (t+1)^2 \Pr[R = t] \qquad (t := k-1)$$

Variance - Examples

Continuing from the previous slide,

$$\mathbb{E}[R^2|A^c] = \sum_{t=0}(t+1)^2 \Pr[R=t] = \sum_{t=0} t^2 \Pr[R=t] + 2\sum_{t=0} t \Pr[R=t] + \sum_{t=0} \Pr[R=t]$$
$$= \sum_{t=1} t^2 \Pr[R=t] + 2\sum_{t=1} t \Pr[R=t] + \sum_{t=1} \Pr[R=t] = \mathbb{E}[R^2] + 2\mathbb{E}[R] + 1$$

Putting everything together,

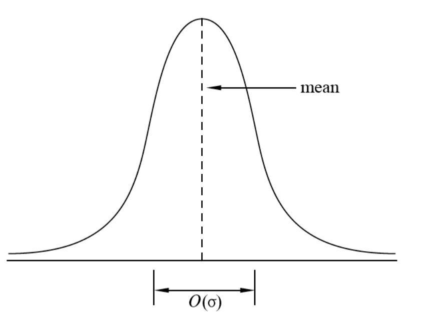
$$\mathbb{E}[R^2] = (1)\,(p) + (\mathbb{E}[R^2] + 2\mathbb{E}[R] + 1])\,(1-p) \implies p\,\mathbb{E}[R^2] = p + 2(1-p)\mathbb{E}[R] + (1-p)\mathbb{E}[1]$$
$$\implies p\,\mathbb{E}[R^2] = p + \frac{2(1-p)}{p} + (1-p) \qquad (\mathbb{E}[R] = \tfrac{1}{p},\ \mathbb{E}[1] = 1)$$
$$\implies \mathbb{E}[R^2] = \frac{2(1-p)}{p^2} + \frac{1}{p} \implies \mathbb{E}[R^2] = \frac{2-p}{p^2}$$
$$\implies \mathrm{Var}[R] = \mathbb{E}[R^2] - (\mathbb{E}[R])^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

Standard Deviation

Standard Deviation: For r.v. X, the standard deviation in X is defined as:

$$\sigma_X := \sqrt{\text{Var}[X]} = \sqrt{\mathbb{E}[X^2] - (\mathbb{E}[X])^2}$$

Standard deviation has the same units as expectation.



Standard deviation for a "bell"-shaped distribution indicates how wide the "main part" of the distribution is.

Properties of Variance

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

Proof:

$$\begin{aligned}
\text{Var}[aR + b] &= \mathbb{E}[(aR + b)^2] - (\mathbb{E}[aR + b])^2 = \mathbb{E}[a^2 R^2 + 2abR + b^2] - (\mathbb{E}[aR] + \mathbb{E}[b])^2 \\
&= (a^2\mathbb{E}[R^2] + 2ab\mathbb{E}[R] + b^2) - (a\mathbb{E}[R] + b)^2 \\
&= (a^2\mathbb{E}[R^2] + 2ab\mathbb{E}[R] + b^2) - (a^2(\mathbb{E}[R])^2 + 2ab\mathbb{E}[R] + b^2) \\
&= a^2 \left[\mathbb{E}[R^2] - (\mathbb{E}[R])^2\right]
\end{aligned}$$

$$\implies \text{Var}[aR + b] = a^2\text{Var}[R]$$

Similarly, for the standard deviation,

$$\sigma_{aR+b} = \sqrt{\text{Var}[aR + b]} = \sqrt{a^2\text{Var}[R]} = |a|\,\sigma_R$$

Note the difference from the property of expectation,

$$\mathbb{E}[aR + b] = a\mathbb{E}[R] + b$$

Properties of Variance

Recall that for r.v's R and S, $\mathbb{E}[R+S] = \mathbb{E}[R] + \mathbb{E}[S]$. In general, such a property is not true for the variance, i.e. variance of a sum is not necessarily equal to the sum of the variances.

If the r.v's R and S are *independent*, $\text{Var}[R+S] = \text{Var}[R] + \text{Var}[S]$.

Proof:

$$\begin{aligned}
\text{Var}[R+S] &= \mathbb{E}[(R+S)^2] - (\mathbb{E}[R+S])^2 = \mathbb{E}[R^2 + S^2 + 2RS] - (\mathbb{E}[R] + \mathbb{E}[S])^2 \\
&= \mathbb{E}[R^2 + S^2 + 2RS] - [(\mathbb{E}[R])^2 + (\mathbb{E}[S])^2 + 2\mathbb{E}[R]\,\mathbb{E}[S]] \\
&= [\mathbb{E}[R^2] - (\mathbb{E}[R])^2] + [\mathbb{E}[S^2] - (\mathbb{E}[S])^2] + 2(\mathbb{E}[RS] - \mathbb{E}[R]\,\mathbb{E}[S]) \\
&= \text{Var}[R] + \text{Var}[S] + 2(\mathbb{E}[RS] - \mathbb{E}[R]\,\mathbb{E}[S])
\end{aligned}$$

Recall that if r.v. are independent, $\mathbb{E}[RS] = \mathbb{E}[R]\,\mathbb{E}[S]$,

$$\implies \text{Var}[R+S] = \text{Var}[R] + \text{Var}[S]$$

Properties of Variance

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 $y \in \text{Range}(R_j)$, events $\Pr[R_i = x]$ and $\Pr[R_j = y]$ are pairwise independent implying that $\Pr[(R_i = x) \cap (R_j = y)] = \Pr[R_i = x] \Pr[R_j = y]$.

We can prove that for any pair of pairwise independent r.v's, R_i and R_j, $\mathbb{E}[R_i R_j] = \mathbb{E}[R_i] \, \mathbb{E}[R_j]$.

For pairwise independent random variables $R_1, R_2, R_3, \ldots R_n$, $\text{Var}[\sum_{i=1}^{n} R_i] = \sum_{i=1}^{n} \text{Var}[R_i]$.

$$\textit{Proof}: \text{Var}[R_1 + R_2 + \ldots R_n] = \mathbb{E}[(R_1 + R_2 + \ldots R_n)^2] - (\mathbb{E}[R_1 + R_2 + \ldots R_n])^2$$

$$= \sum_{i=1}^{n} [\mathbb{E}[R_i^2] - (\mathbb{E}[R_i])^2] + 2 \sum_{i,j \mid 1 \le i < j \le n} [\mathbb{E}[R_i R_j] - \mathbb{E}[R_i] \, \mathbb{E}[R_j]]$$

$$\implies \text{Var}[R_1 + R_2 + \ldots R_n] = \sum_{i=1}^{n} \text{Var}[R_i] \qquad \text{(Since the r.v's are pairwise independent)}$$

Importantly, we do not require the r.v's to be mutually independent. Mutual independence $\implies$ pairwise independence, but pairwise independence $\not\Rightarrow$ mutual independence.

Variance - Examples

Q: If $R \sim \text{Bin}(n, p)$, calculate $\text{Var}[R]$.

Define R_i to be the indicator random variable that we get a heads in toss i of the coin. Recall that R is the random variable equal to the number of heads in n tosses.

Hence,

$$R = R_1 + R_2 + \ldots + R_n \implies \text{Var}[R] = \text{Var}[R_1 + R_2 + \ldots + R_n]$$

Since $R_1, R_2, \ldots, R_n$ are mutually independent indicator random variables,

$$\text{Var}[R] = \text{Var}[R_1] + \text{Var}[R_2] + \ldots + \text{Var}[R_n]$$

Since the variance of an indicator (Bernoulli) r.v. is $p(1-p)$,

$$\text{Var}[R] = n\,p\,(1-p).$$

Questions?

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).

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

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

(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 75 students, on average, there are $\frac{(75)\,(37)}{365} = 7.60$ 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 75 students, the standard deviation for the matching birthdays is equal to $\sqrt{\frac{(37)(75)}{365}\frac{364}{365}} \approx 2.75$.

Questions?