CMPT 210: Probability and Computing

Lecture 19

Sharan Vaswani

November 14, 2024

Recap: Randomized Quick Select

Aim: Given an array A of n distinct numbers, return the k^{th} smallest element in A for $k \in [1, n]$.

Algorithm Randomized Quick Select

---

function QuickSelect(A, k)
If Length(A) = 1, return A[1].
Select $p \in A$ uniformly at random.
Construct sets Left $:= \{x \in A | x < p\}$ and Right $:= \{x \in A | x > p\}$.
$r = |\text{Left}| + 1$ {Element p is the r^{th} smallest element in A.}
if $k = r$ **then**
    return p
else if $k < r$ **then**
    QuickSelect(Left, k)
else
    QuickSelect(Right, $k - r$)
end if

Recap: Randomized Quick Select – Analysis

- In the worst case, Randomized Quick Select has an $O(n^2)$ runtime which is worse than the naive strategy of sorting and returning the k^{th} element.
- **Claim**: For any array A with n distinct elements, and for any $k \in [n]$, Randomized Quick Select performs fewer than $8n$ comparisons in expectation.
- Last time, we proved that the child sub-problem's array (either Left or Right) after the partitioning (in Line 4 of the algorithm) has expected size smaller than $\frac{7n}{8}$.

Randomized Quick Select – Analysis

In order to upper-bound the total number of comparisons, we use the Lemma with a strong induction on n. Recall that we need to prove that Randomized Quick Select requires fewer than $8n$ comparisons in expectation.

Base case: If $n = 1$, then we require $0 < 8(1)$ comparisons. Hence the base case is satisfied.

Inductive Step: Assume that for all $m < n$,
$\mathbb{E}[\text{Total number of comparisons for size } m \text{ array}] < 8\,m$.

$$
\begin{aligned}
&\mathbb{E}[\text{Total number of comparisons for size } n \text{ array}] \\
&= \mathbb{E}[(n-1) + \text{Total number of comparisons in child sub-problem}] && \text{(First step of algorithm)} \\
&= (n-1) + \mathbb{E}[\text{Total number of comparisons in child sub-problem}] && \text{(Linearity of expectation)} \\
&< (n-1) + 8\,\mathbb{E}[|\text{Child}|] && \text{(Induction hypothesis)} \\
&< (n-1) + 8\frac{7n}{8} < 8n. && \text{(Lemma)}
\end{aligned}
$$

- Hence, for any $k \in [n]$, on average, Randomized Quick Select requires fewer than $8n$ comparisons, even though it might require $O(n^2)$ comparisons in the worst-case.

Questions?

Deviation from the Mean

- We have developed tools to calculate the mean of random variables. Getting a handle on the expectation is useful because it tell us what would happen on average.

- However, summarizing the PDF using the mean is typically not enough. We also want to know how "spread" the distribution is.

Example: Consider three random variables W, Y and Z whose PDF's can be given as:

$$\begin{aligned} W &= 0 && \text{(with } p = 1) \\ Y &= -1 && \text{(with } p = 1/2) \\ &= +1 && \text{(with } p = 1/2) \\ Z &= -1000 && \text{(with } p = 1/2) \\ &= +1000 && \text{(with } p = 1/2) \end{aligned}$$

Though $\mathbb{E}[W] = \mathbb{E}[Y] = \mathbb{E}[Z] = 0$, these distributions are quite different. Z can take values really far away from its expected value, while W can take only one value equal to the mean. Hence, we want to understand how much does a random variable "deviate" from its mean.

Deviation from the Mean

- Before we calculate the deviation of a r.v. from its mean, we need an additional definition.
- For a r.v. $X : \mathcal{S} \to V$ and a function $g : V \to \mathbb{R}$, we define $\mathbb{E}[g(X)]$ as follows:

$$\mathbb{E}[g(X)] := \sum_{x \in \text{Range}(X)} g(x) \Pr[X = x]$$

If $g(x) = x$ for all $x \in \text{Range}(X)$, then $\mathbb{E}[g(X)] = \mathbb{E}[X]$.

Q: For a standard dice, if X is the r.v. corresponding to the number that comes up on the dice, compute $\mathbb{E}[X^2]$ and $(\mathbb{E}[X])^2$

For a standard dice, $X \sim \text{Uniform}(\{1, 2, 3, 4, 5, 6\})$ and hence,

$$\mathbb{E}[X^2] = \sum_{x \in \{1,2,3,4,5,6\}} x^2 \Pr[X = x] = \frac{1}{6}\left[1^2 + 2^2 + \ldots + 6^2\right] = \frac{91}{6}$$

$$(\mathbb{E}[X])^2 = \left(\sum_{x \in \{1,2,3,4,5,6\}} x \Pr[X = x]\right)^2 = \left(\frac{1}{6}\left[1 + 2 + \ldots + 6\right]\right)^2 = \frac{49}{4}$$

Variance

Definition: *Variance* is the standard way to measure the deviation of a r.v. from its mean. Formally, for a r.v. X,

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in \mathrm{Range}(X)} (x - \mu)^2 \Pr[X = x] \qquad \text{(where } \mu := \mathbb{E}[X])$$

Intuitively, the variance measures the weighted (by the probability) average of how far (in squared distance) the random variable is from its mean μ.

Q: If $X \sim \mathrm{Ber}(p)$, compute $\mathrm{Var}[X]$.

Since X is a Bernoulli random variable, $X = 1$ with probability p and $X = 0$ with probability $1 - p$. Recall that $\mathbb{E}[X] = \mu = (0)(1 - p) + (1)(p) = p$.

$$\mathrm{Var}[X] = \sum_{x \in \{0,1\}} (x - p)^2 \Pr[X = x] = (0 - p)^2 \Pr[X = 0] + (1 - p)^2 \Pr[X = 1]$$

$$= p^2(1 - p) + (1 - p)^2 p = p(1 - p)[p + 1 - p] = p(1 - p).$$

- For a Bernoulli r.v. X, $\mathrm{Var}[X] = p(1 - p) \leq \frac{1}{4}$ and the variance is maximum when $p = 1/2$.

Variance

Alternate definition of variance: $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mu^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Proof : $\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in \mathrm{Range}(X)} (x - \mu)^2 \Pr[X = x]$

$$= \sum_{x \in \mathrm{Range}(X)} (x^2 - 2\mu x + \mu^2) \Pr[X = x]$$

$$= \sum_{x \in \mathrm{Range}(X)} (x^2 \Pr[X = x]) - (2\mu x \Pr[X = x]) + (\mu^2) \Pr[X = x]$$

$$= \sum_{x \in \mathrm{Range}(X)} x^2 \Pr[X = x] - 2\mu \sum_{x \in \mathrm{Range}(X)} x \Pr[X = x] + \mu^2 \sum_{x \in \mathrm{Range}(X)} \Pr[X = x]$$

(Since μ is a constant does not depend on the x in the sum.)

$$= \mathbb{E}[X^2] - 2\mu\,\mathbb{E}[X] + \mu^2 \sum_{x \in \mathrm{Range}(X)} \Pr[X = x] \quad \text{(Definition of } \mathbb{E}[X] \text{ and } \mathbb{E}[X^2])$$

$$= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 \quad \text{(Definition of } \mu)$$

$$\implies \mathrm{Var}[X] = \mathbb{E}[X^2] - \mu^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Back to throwing dice

Q: For a standard dice, if X is the r.v. equal to the number that comes up, compute $\text{Var}[X]$.

Recall that, for a standard dice, $X \sim \text{Uniform}(\{1, 2, 3, 4, 5, 6\})$ and hence,

$$\mathbb{E}[X^2] = \sum_{x \in \{1,2,3,4,5,6\}} x^2 \Pr[X = x] = \frac{1}{6}\left[1^2 + 2^2 + \ldots + 6^2\right] = \frac{91}{6}$$

$$(\mathbb{E}[X])^2 = \left(\sum_{x \in \{1,2,3,4,5,6\}} x \Pr[X = x]\right)^2 = \left(\frac{1}{6}[1 + 2 + \ldots + 6]\right)^2 = \frac{49}{4}$$

$$\implies \text{Var}[X] = \frac{91}{6} - \frac{49}{4} \approx 2.917$$

Q: If $X \sim \text{Uniform}(\{v_1, v_2, \ldots v_n\})$, compute $\text{Var}[X]$.

$$\mathbb{E}[X] = \sum_{i=1}^{n} v_i \Pr[X = v_i] = \frac{1}{n}[v_1 + v_2 + \ldots v_n] \quad ; \quad \mathbb{E}[X^2] = \frac{1}{n}[v_1^2 + v_2^2 + \ldots v_n^2].$$

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

Variance - Examples

Q: Calculate Var[W], Var[Y] and Var[Z] whose PDF's are given as:

$$
\begin{aligned}
W &= 0 && \text{(with } p = 1) \\
Y &= -1 && \text{(with } p = 1/2) \\
&= +1 && \text{(with } p = 1/2) \\
Z &= -1000 && \text{(with } p = 1/2) \\
&= +1000 && \text{(with } p = 1/2)
\end{aligned}
$$

Recall that $\mathbb{E}[W] = \mathbb{E}[Y] = \mathbb{E}[Z] = 0$.

$\text{Var}[W] = \mathbb{E}[W^2] - (\mathbb{E}[W])^2 = \mathbb{E}[W^2] = \sum_{w \in \text{Range}(W)} w^2 \Pr[W = w] = 0^2(1) = 0$. The variance of W is zero because it can only take one value and the r.v. does not "vary".

$\text{Var}[Y] = \mathbb{E}[Y^2] = \sum_{y \in \text{Range}(Y)} y^2 \Pr[Y = y] = (-1)^2(1/2) + (1)^2(1/2) = 1$.

$\text{Var}[Z] = \mathbb{E}[Z^2] = \sum_{z \in \text{Range}(Z)} z^2 \Pr[Z = z] = (-1000)^2(1/2) + (1000)^2(1/2) = 10^6$.

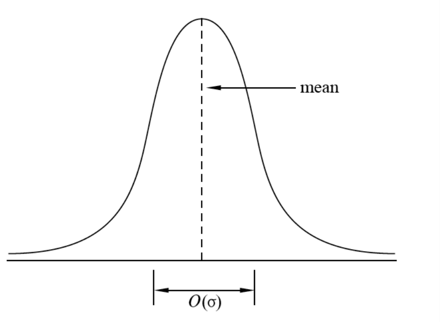
- Hence, the variance can be used to distinguish between r.v.'s that have the same mean.

Standard Deviation

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

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

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.