CMPT 210: Probability and Computing

Lecture 17

Sharan Vaswani

March 10, 2023

Recap

Two random variables R_1 and R_2 are independent if for *all* $x_1 \in \text{Range}(R_1)$ and $x_2 \in \text{Range}(R_2)$, events $[R_1 = x_1]$ and $[R_2 = x_2]$ are independent. More formally, we require that for *all* $x_1 \in \text{Range}(R_1)$ and $x_2 \in \text{Range}(R_2)$,

$$\Pr[(R_1 = x_1) \cap (R_2 = x_2)] = \Pr[(R_1 = x_1)] \Pr[(R_2 = x_2)].$$

Independence of random variables

Q: If R_1 and R_2 are not independent, is $\mathbb{E}[R_1 + R_2] = \mathbb{E}[R_1] + \mathbb{E}[R_2]$?

Yes! Recall the derivation of the linearity of expectation. We never assumed that R_1 and R_2 are independent for the proof and the linearity of expectation holds regardless of whether the random variables are independent.

Q: If R_1 and R_2 are independent, is $\mathbb{E}[R_1 R_2] = \mathbb{E}[R_1]\,\mathbb{E}[R_2]$? Yes!

$$
\begin{aligned}
\mathbb{E}[R_1 R_2] &= \sum_{x \in \text{Range}(R_1 R_2)} x \Pr[R_1 R_2 = x] = \sum_{r_1 \in \text{Range}(R_1), r_2 \in \text{Range}(R_2)} r_1 r_2 \Pr[R_1 = r_1 \cap R_2 = r_2] && (x = r_1\, r_2) \\
&= \sum_{r_1 \in \text{Range}(R_1)} \sum_{r_2 \in \text{Range}(R_2)} r_1 r_2 \Pr[R_1 = r_1 \cap R_2 = r_2] && \text{(Splitting the sum)} \\
&= \sum_{r_1 \in \text{Range}(R_1)} \sum_{r_2 \in \text{Range}(R_2)} r_1 r_2 \Pr[R_1 = r_1] \Pr[R_2 = r_2] && \text{(Independence)} \\
&= \sum_{r_1 \in \text{Range}(R_1)} r_1 \Pr[R_1 = r_1] \sum_{r_2 \in \text{Range}(R_2)} r_2 \Pr[R_2 = r_2] = \mathbb{E}[R_1]\mathbb{E}[R_2]
\end{aligned}
$$

Independence of random variables

Alternate definition of independence – two random variables R_1 and R_2 are independent if for *all* $x_1 \in \text{Range}(R_1)$ and $x_2 \in \text{Range}(R_2)$,

$$\Pr[(R_1 = x_1)|(R_2 = x_2)] = \Pr[(R_1 = x_1)]$$
$$\Pr[(R_2 = x_2)|(R_1 = x_1)] = \Pr[(R_2 = x_2)]$$

Similar to events, random variables $R_1, R_2, \ldots, R_n$ are mutually independent if for all $x_1, x_2, \ldots, x_n$, events $[R_1 = x_1], [R_2 = x_2], \ldots [R_n = x_n]$ are mutually independent.

Mutual Independence of events: A set of events is said to be mutually independent if the probability of each event in the set is the same no matter which of the events has occurred. For events E_1, E_2 and E_3 to be mutually independent, all the following equalities should hold:

$$\Pr[E_1 \cap E_2] = \Pr[E_1] \Pr[E_2] \quad \Pr[E_1 \cap E_3] = \Pr[E_1] \Pr[E_3]$$
$$\Pr[E_2 \cap E_3] = \Pr[E_2] \Pr[E_3] \quad \Pr[E_1 \cap E_2 \cap E_3] = \Pr[E_1] \Pr[E_2] \Pr[E_3].$$

Alternatively, (i) $\forall i$ and $j \neq i$, $\Pr[E_i|E_j] = \Pr[E_i]$ and (ii) $\forall i$ and $j, k \neq i$, $\Pr[E_i|E_j \cap E_k] = \Pr[E_i]$.

Expectation/Independence - Examples

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 $\frac{1}{n}$. What is the expected number of people who get their own coat?

Let G be the number of people that get their own coat. We wish to compute $\mathbb{E}[G]$. Define G_i to be the indicator r.v. that person i gets their own coat. Observe that $G = G_1 + G_2 + \ldots + G_n$ and by linearity of expectation $\mathbb{E}[G] = \mathbb{E}[G_1] + \mathbb{E}[G_2] + \ldots + \mathbb{E}[G_n]$. For each i, $\mathbb{E}[G_i] = \Pr[G_i] = \frac{1}{n}$. Hence, $\mathbb{E}[G] = 1$ meaning that on average one person will correctly receive their coat.

Q: If G_i is the indicator r.v. that person i gets their own coat, are the random variables $G_1, G_2, \ldots G_n$ mutually independent?

No. Since if $G_1 = G_2 = \ldots G_{n-1} = 1$, then, $\Pr[G_n = 1|(G_1 = 1 \cap G_2 = 1 \cap \ldots \cap G_{n-1} = 1)] = 1 \neq \frac{1}{n} = \Pr[G_n = 1]$. Conditioning on $(G_1, G_2, \ldots, G_{n-1})$ changes $\Pr[G_n]$, and hence the r.v's are not independent. Notice that we have used the linearity of expectation even though these r.v's are not mutually independent.

Questions?

Joint distributions

For a given experiment, we are often interested not only in the PDFs of individual random variables but also in the relationships between two or more random variables. For example, we might be interested in the mean time of failure and its connection with different number of components in the system.

A joint distribution between r.v's X and Y can be specified by its joint PDF as follows:

$$\mathrm{PDF}_{X,Y}[x,y] = \Pr[X = x \cap Y = y]$$

If X and Y are independent random variables, $\mathrm{PDF}_{X,Y}[x,y] = \mathrm{PDF}_X[x]\,\mathrm{PDF}_Y[y]$.

If $\mathrm{Range}[X] = \{x_1, x_2, \ldots x_n\}$, $\mathrm{Range}[Y] = \{y_1, y_2, \ldots y_n\}$, then for $x \in \mathrm{Range}(X)$,
$[X = x] = [X = x \cap y = y_1] \cup [X = x \cap y = y_2] \cup \ldots \cup [X = x \cap y = y_n]$
$\implies \Pr[X = x] = \Pr[X = x \cap y = y_1] + \Pr[X = x \cap y = y_2] + \ldots + \Pr[X = x \cap y = y_n]$.

$\implies \mathrm{PDF}_X[x] = \sum_i \mathrm{PDF}_{X,Y}[x, y_i]$.
Hence, we can obtain the distribution for each r.v. from the joint distribution by "marginalizing" over the other r.v's.

Joint distributions - Examples

Q: Suppose that 3 batteries are randomly chosen from a group of 3 new, 4 used but still working, and 5 defective batteries. If we let X and Y denote, respectively, the number of new and used but still working batteries that are chosen, completely specify $\mathrm{PDF}_{X,Y}$.

For $i \in [3], j \in [3]$, $\mathrm{PDF}_{X,Y}[i,j] = \Pr[X = i \cap Y = j \mid X + Y \leq 3] = \frac{\binom{3}{i}\binom{4}{j}\binom{5}{3-i-j}}{\binom{12}{3}}$.

$\mathrm{PDF}_{X,Y}[0,0] = \frac{\binom{5}{3}}{\binom{12}{3}} = 10/220$, $\mathrm{PDF}_{X,Y}[1,2] = \frac{\binom{3}{1}\binom{4}{2}\binom{5}{2}}{\binom{12}{3}} = 18/220$.

Table 4.1 $P\{X = i, Y = j\}$.

i \ j	**0**	**1**	**2**	**3**	**Row Sum** $= P\{X = i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Column Sums $= P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

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.

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:

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

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.

Variance

Standard way to measure the deviation from the mean is to calculate the *variance*. For 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 the 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, $\mathrm{Var}[X] = p(1-p) \leq \frac{1}{4}$. Hence, the variance is maximum when $p = 1/2$ (equal probability of getting heads/tails).

Variance

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

$$\begin{aligned}
\textit{Proof}: \text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] &= \sum_{x \in \text{Range}(X)} (x-\mu)^2 \Pr[X = x] \\
&= \sum_{x \in \text{Range}(X)} (x^2 - 2\mu x + \mu^2) \Pr[X = x] \\
&= \sum_{x \in \text{Range}(X)} (x^2 \Pr[X = x]) - (2\mu x \Pr[X = x]) + (\mu^2) \Pr[X = x] \\
&= \sum_{x \in \text{Range}(X)} x^2 \Pr[X = x] - 2\mu \sum_{x \in \text{Range}(X)} x \Pr[X = x] + \mu^2 \sum_{x \in \text{Range}(X)} \Pr[X = x] \\
&\qquad \text{(Since } \mu \text{ is a constant does not depend on the } x \text{ in the sum.)} \\
&= \mathbb{E}[X^2] - 2\mu \mathbb{E}[X] + \mu^2 \sum_{x \in \text{Range}(X)} \Pr[X = x] \quad \text{(Definition of } \mathbb{E}[X] \text{ and } \mathbb{E}[X^2]\text{)} \\
&= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 \qquad \text{(Definition of } \mu\text{)} \\
\implies \text{Var}[X] &= \mathbb{E}[X^2] - \mu^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.
\end{aligned}$$

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,

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

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

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

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

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

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

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.

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

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

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,

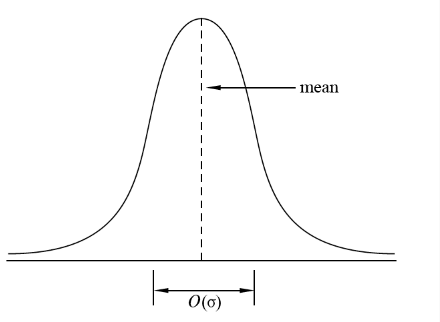
$$\mathbb{E}[R^2] = (1)\,(p) + (\mathbb{E}[R^2] + 2\mathbb{E}[R] + 1])\,(1-p) \implies p\,\mathbb{E}[R^2] = p + 2(1-p)\mathbb{E}[R] + (1-p)\mathbb{E}[1]$$
$$\implies p\,\mathbb{E}[R^2] = p + \frac{2(1-p)}{p} + (1-p) \qquad (\mathbb{E}[R] = \tfrac{1}{p},\ \mathbb{E}[1] = 1)$$
$$\implies \mathbb{E}[R^2]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

Standard Deviation

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

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

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.

Properties of Variance

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

Proof:

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

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

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*, $\mathrm{Var}[R+S] = \mathrm{Var}[R] + \mathrm{Var}[S]$.

Proof:

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

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

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

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

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

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

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

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?