# CMPT 210: Probability and Computation

Lecture 15

Sharan Vaswani

July 5, 2022

# Logistics

- Collect your Midterm exams from TASC-1 9203 on Tuesdays between 10.30 am - 12 pm.
- Assignment 3 is out: `https://vaswanis.github.io/210-S22/A3.pdf`
  Due Friday 15 July in class.
- For A3, you can use your late-submission and submit on Tuesday 19 July in class.
- Solutions will be released on 19 July after class, meaning that no submissions will be allowed after that.
- If you have used your late-submission, and submit late again, you will lose 50% of the mark.
- If you have questions about either assignment or the marking, post it on Piazza:
  `https://piazza.com/sfu.ca/summer2022/cmpt210/home`

# Recap

**Expectation**/mean of a random variable $R$ is denoted by $\mathbb{E}[R]$ and "summarizes" its distribution. Formally, $\mathbb{E}[R] := \sum_{\omega \in \mathcal{S}} \Pr[\omega]\, R[\omega]$

**Alternate definition of expectation**: $\mathbb{E}[R] = \sum_{x \in \text{Range}(R)} x \Pr[R = x]$.

**Linearity of Expectation**: For $n$ random variables $R_1, R_2, \ldots, R_n$ and constants $a_1, a_2, \ldots, a_n$, $\mathbb{E}\left[\sum_{i=1}^n a_i R_i\right] = \sum_{i=1}^n a_i\, \mathbb{E}[R_i]$.

**Conditional Expectation**: For random variable $R$, the expected value of $R$ conditioned on an event A is given by:

$$\mathbb{E}[R|A] = \sum_{x \in \text{Range}(R)} x \Pr[R = x|A]$$

**Law of Total Expectation**: If $R$ is a random variable $\mathcal{S} \to V$ and events $A_1, A_2, \ldots A_n$ form a partition of the sample space, then,

$$\mathbb{E}[R] = \sum_i \mathbb{E}[R|A_i] \Pr[A_i]$$

# Independence of random variables

We define two random variables $R_1$ and $R_2$ to be 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,

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

Q: Suppose we toss three independent, unbiased coins. Let $C$ be r.v. equal to the number of heads that appear and $M$ be the r.v. that is equal to 1 if all the coins match. Are random variables $C$ and $M$ independent?

$\text{Range}(C) = \{0, 1, 2, 3\}$ and $\text{Range}(M) = \{0, 1\}$. $\Pr[C = 3] = \frac{1}{8}$ and $\Pr[M = 1] = \frac{1}{4}$. $\Pr[(C = 3) \cap (M = 1)] = \frac{1}{8} \neq \frac{1}{32} = \Pr[C = 3] \Pr[M = 1]$. Hence, $C$ and $M$ are not independent.

# Independence - Examples

Q: If $H_1$ is the indicator r.v. equal to one if the first toss is a heads, are $H_1$ and $M$ independent?
$\Pr[H_1 = 1] = \Pr[H_1 = 0] = \frac{1}{2}$, $\Pr[M = 1] = \frac{1}{4}$, $\Pr[M = 0] = \frac{3}{4}$.
$\Pr[H_1 = 0 \cap M = 1] = \Pr[\{TTT\}] = \frac{1}{8} = \Pr[H_1 = 0] \Pr[M = 1]$.
$\Pr[H_1 = 1 \cap M = 1] = \Pr[\{HHH\}] = \frac{1}{8} = \Pr[H_1 = 1] \Pr[M = 1]$.
$\Pr[H_1 = 0 \cap M = 0] = \Pr[\{THH, THT, TTH\}] = \frac{3}{8} = \Pr[H_1 = 0] \Pr[M = 0]$.
$\Pr[H_1 = 1 \cap M = 0] = \Pr[\{HHT, HTH, HTT\}] = \frac{3}{8} = \Pr[H_1 = 1] \Pr[M = 0]$.
Hence, $H_1$ and $M$ are independent.

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.

# Independence - Examples

Q: Suppose that the successive daily changes of the price of a given stock are assumed to be independent and identically distributed random variables – for each day $i$, the PDF is:

$$\begin{aligned}
\Pr[\text{Daily change on day } i] &= -3 && \text{(With } p = 0.1,\text{)} \\
&= -2 && \text{(With } p = 0.1\text{)} \\
&= -1 && \text{(With } p = 0.2\text{)} \\
&= 0 && \text{(With } p = 0.3\text{)} \\
&= 1 && \text{(With } p = 0.2\text{)} \\
&= 2 && \text{(With } p = 0.1\text{)}
\end{aligned}$$

If $E$ is the event that the stocks price will increase successively by 1, 2, and 0 points in the next three days, compute $\mathbb{E}[\mathcal{I}_E]$.

If $X_i$ is the r.v. corresponding to the price increase on day $i$, we wish to compute $\mathbb{E}[\mathcal{I}_E] = \Pr[E] = \Pr[X_1 = 1 \cap X_2 = 2 \cap X_3 = 0]$. $X_1$, $X_2$ and $X_3$ are mutually independent and hence, $\Pr[X_1 = 1 \cap X_2 = 2 \cap X_3 = 0] = \Pr[X_1 = 1] \Pr[X_2 = 2] \Pr[X_3 = 0] = 0.006$.

# 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!

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

# 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 $x_2 \in \text{Range}(R_2)$,

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

Intuitively, this means that conditioning on the value of $R_2$ does not change the probability of the event $R_1 = x_1$, and vice-versa.

# Expectation - 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,
$\Pr[G_n = 1 | (G_1 = 1 \cap G_2 = 1 \cap \ldots \cap G_{n-1} = 1)] = 1 \neq \frac{1}{n} = \Pr[G_n = 1]$. Notice that we have used the linearity of expectation for the $G_i$'s even though these r.v. are not mutually independent.

# Expectation - Examples

For a random variable $X : \mathcal{S} \to V$ and a function $g : V \to \mathbb{R}$, we define $\mathbb{E}[g(X)]$ as follows:

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

If $g(x) = x$ for all $x \in \mathrm{Range}(X)$, 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, $X \sim \mathrm{Uniform}(\{1, 2, 3, 4, 5, 6\})$ and hence,

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

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

**Questions?**

# Joint distribution

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 distribution - 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 $i \in [3], j \in [3]$,

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

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

**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 the random variable is from the mean $\mu$.

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 $\mathbb{E}[X] = \mu = (0)(1-p) + (1)(p) = p$.

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

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

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

# Back to throwing dice

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

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

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

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

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

In general, if $X \sim \mathrm{Uniform}(\{v_1, v_2, \ldots v_n\})$,

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

# 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".

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

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

The variance of $Z$ is the largest because it can take values that are far away from the mean. Hence, the variance can be used to distinguish between r.v.'s that have the same mean.

# Variance - Examples

Q: Game A: We win \$2 with probability 2/3 and lose \$1 with probability 1/3. Game B: We win \$1002 with probability 2/3 and lose \$2001 with probability 1/3. Which game is better financially? $\mathbb{E}[A] = 2\left(\frac{2}{3}\right) - 1\left(\frac{1}{3}\right) = \$1$. Similarly, $\mathbb{E}[B] = 1002\left(\frac{2}{3}\right) - 2001\left(\frac{1}{3}\right) = \$1$.

Hence, on average, the two games have the same payoff. To get more information, let us analyze the variance. $\text{Var}[A] = \mathbb{E}[A^2] - 1 = 2^2\left(\frac{2}{3}\right) + (-1)^2\left(\frac{1}{3}\right) - 1 = 2$. Similarly, $\text{Var}[B] = 2004002$.

Intuitively, this means that the payoff in Game A is usually close to the expected value of \$1, but the payoff in Game B can deviate very far from this expected value. High variance is often associated with high risk. For example, in ten rounds of Game A, we expect to make \$10, but could conceivably lose \$10 instead (if we lose each game). On the other hand, in ten rounds of game B, we also expect to make \$10, but could actually lose more than \$20000!

# Variance - Examples

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

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

Similar to Slide 13 of Lecture 14, let $A$ be the event that we get a heads in the first toss. Using the law of total expectation,

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

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

$$\mathbb{E}[R^2] = (1)(p) + \mathbb{E}[R^2|A^c](1-p)$$

$$\mathbb{E}[R^2|A^c] = \sum_{k=1} k^2 \Pr[R = k|A^c]$$

$$\Pr[R = k| \text{ if first toss is a tails}] = \Pr[R = k-1]$$

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

# Variance - Examples

Putting everything together,

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

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

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

$$\implies \mathrm{Var}[R] = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

**Questions?**