CMPT 210: Probability and Computing

Lecture 20

Sharan Vaswani

March 26, 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: $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

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

Covariance

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

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

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

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.

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

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 $\mathrm{Cov}[R,S] = 0$.

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,

$$\mathrm{Var}[R+S] = \underbrace{\mathrm{Var}[R]}_{\text{How much R varies}} + \underbrace{\mathrm{Var}[S]}_{\text{How much S varies}} + \underbrace{2\,\mathrm{Cov}[R,S]}_{\text{How much both of them vary together}}$$

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 19, Slide 6, 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.

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 19, Slide 6, 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.

We can do covariance for more than two random 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 $\mathrm{Var}[2R] = 2^2\mathrm{Var}[R]$.

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 19, Slide 6, 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 19, Slide 7):

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

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

Cov(X, Y) = E[XY] - E[X]E[Y]

E[XY] - Pr(I_A)Pr(I_B)

XY = 1 if X = Y = 1, both events happen

XY is the indicator for both events to happen.

E[XY] = Pr(I_A = 1 and I_B = 1)

Cov(X, Y) = Pr(I_A = 1 and I_B = 1) - Pr(I_A)Pr(I_B)

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 $\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$. 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.

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 $\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$. 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.

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

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

If $\text{Cov}[X, Y] > 0 \implies \Pr[A \cap B] > \Pr[A]\Pr[B]$. 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

Correlation is normalized, in between [-1, 1].

The correlation between two r.v's R_1 and R_2 is defined as:

The sign of the correlation is important

$$\mathrm{Corr}[R_1, R_2] = \frac{\mathrm{Cov}[R_1, R_2]}{\sqrt{\mathrm{Var}[R_1]\,\mathrm{Var}[R_2]}}$$

$\mathrm{Corr}[R_1, R_2] \in [-1, 1]$ and indicates the strength of the relationship between R_1 and R_2.

If R1 and R2 are independent, then the Corr(R1, R2) = 0.

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.

Correlation

The correlation between two r.v's R_1 and R_2 is defined as:

$$\text{Corr}[R_1, R_2] = \frac{\text{Cov}[R_1, R_2]}{\sqrt{\text{Var}[R_1]\,\text{Var}[R_2]}}$$

$\text{Corr}[R_1, R_2] \in [-1, 1]$ and indicates the strength of the relationship between R_1 and R_2.

If $\text{Corr}[R_1, R_2] > 0$, then R_1 and R_2 are said to be positively correlated, else if $\text{Corr}[R_1, R_2] < 0$, the r.v's are negatively correlated.

If $R_1 = R_2 = R$, then, $\text{Corr}[R, R] = \frac{\text{Cov}[R,R]}{\sqrt{\text{Var}[R]\,\text{Var}[R]}} = \frac{\text{Var}[R]}{\text{Var}[R]} = 1$.

Correlation

The correlation between two r.v's R_1 and R_2 is defined as:

$$\text{Corr}[R_1, R_2] = \frac{\text{Cov}[R_1, R_2]}{\sqrt{\text{Var}[R_1]\,\text{Var}[R_2]}}$$

$\text{Corr}[R_1, R_2] \in [-1, 1]$ and indicates the strength of the relationship between R_1 and R_2.

If $\text{Corr}[R_1, R_2] > 0$, then R_1 and R_2 are said to be positively correlated, else if $\text{Corr}[R_1, R_2] < 0$, the r.v's are negatively correlated.

If $R_1 = R_2 = R$, then, $\text{Corr}[R, R] = \frac{\text{Cov}[R,R]}{\sqrt{\text{Var}[R]\,\text{Var}[R]}} = \frac{\text{Var}[R]}{\text{Var}[R]} = 1$.

If R_1 and R_2 are independent, $\text{Cov}[R_1, R_2] = 0$ and $\text{Corr}[R_1, R_2] = 0$.

Correlation

The correlation between two r.v's R_1 and R_2 is defined as:

$$\text{Corr}[R_1, R_2] = \frac{\text{Cov}[R_1, R_2]}{\sqrt{\text{Var}[R_1]\,\text{Var}[R_2]}}$$

$\text{Corr}[R_1, R_2] \in [-1, 1]$ and indicates the strength of the relationship between R_1 and R_2.

If $\text{Corr}[R_1, R_2] > 0$, then R_1 and R_2 are said to be positively correlated, else if $\text{Corr}[R_1, R_2] < 0$, the r.v's are negatively correlated.

If $R_1 = R_2 = R$, then, $\text{Corr}[R, R] = \frac{\text{Cov}[R,R]}{\sqrt{\text{Var}[R]\,\text{Var}[R]}} = \frac{\text{Var}[R]}{\text{Var}[R]} = 1$.

If R_1 and R_2 are independent, $\text{Cov}[R_1, R_2] = 0$ and $\text{Corr}[R_1, R_2] = 0$.

If $R_1 = -R_2 = R$, then,

$$\begin{aligned}\text{Corr}[R, -R] &= \frac{\text{Cov}[R, -R]}{\sqrt{\text{Var}[R]\,\text{Var}[-R]}} = \frac{\text{Cov}[R, -R]}{\sqrt{\text{Var}[R]\,(-1)^2\text{Var}[R]}} = \frac{\text{Cov}[R, -R]}{\text{Var}[R]} \\ &= \frac{\mathbb{E}[-R^2] - \mathbb{E}[R]\,\mathbb{E}[-R]}{\text{Var}[R]} = \frac{-\mathbb{E}[R^2] + \mathbb{E}[R]\,\mathbb{E}[R]}{\text{Var}[R]} = \frac{-\text{Var}[R]}{\text{Var}[R]} = -1\end{aligned}$$

Questions?

Tail inequalities

Variance gives us one way to measure how "spread" the distribution is.

We want to know the probability that the r.v is smaller or bigger than some number.
By using the tail inequalities, we can measure the probability in the tail of a distribution.

We might wish to measure $Pr(X > 5E[X])$. These type of probabilities are extremely small.

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.

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 $\Pr[X \geq 300] \leq \frac{1}{3}$.

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 $\Pr[X \geq 300] \leq \frac{1}{3}$.

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

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 $x > 0$,

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

if x = cE[X],
then Pr(X > cE[X]) <= 1/c

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 $\mathcal{I}\{X \geq x\}$ to be the indicator r.v. for the event $[X \geq x]$. Then for all values of X, $x\mathcal{I}\{X \geq x\} \leq X$.

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 $\mathcal{I}\{X \geq x\}$ to be the indicator r.v. for the event $[X \geq x]$. Then for all values of X, $x\,\mathcal{I}\{X \geq x\} \leq X$.

$$\begin{aligned} \mathbb{E}[x\,\mathcal{I}\{X \geq x\}] \leq \mathbb{E}[X] \implies x\,\mathbb{E}[\mathcal{I}\{X \geq x\}] \leq \mathbb{E}[X] \implies x\Pr[X \geq x] \leq \mathbb{E}[X] \\ \implies \Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}. \end{aligned}$$

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 $\mathcal{I}\{X \geq x\}$ to be the indicator r.v. for the event $[X \geq x]$. Then for all values of X, $x\,\mathcal{I}\{X \geq x\} \leq X$.

$$\mathbb{E}[x\,\mathcal{I}\{X \geq x\}] \leq \mathbb{E}[X] \implies x\,\mathbb{E}[\mathcal{I}\{X \geq x\}] \leq \mathbb{E}[X] \implies x\Pr[X \geq x] \leq \mathbb{E}[X]$$
$$\implies \Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}.$$

Since the above theorem holds for all $x > 0$, we can set $x = c\mathbb{E}[X]$ for $c \geq 1$. In this case, $\Pr[X \geq c\mathbb{E}[X]] \leq \frac{1}{c}$. Hence, the probability that X is "far" from the mean in terms of the multiplicative factor c is upper-bounded by $\frac{1}{c}$.

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, a person gets their own coat with probability $\frac{1}{n}$.

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, a 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 more than 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.

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.

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 we are provided additional information that X can not take values less than 100 and $\mathbb{E}[X] = 150$, compute the probability that X is at least 200.

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 we are provided additional information 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.

$$\Pr[X \geq 200] = \Pr[Y + 100 \geq 200] = \Pr[Y \geq 100] \leq \frac{\mathbb{E}[Y]}{100} = \frac{50}{100} = \frac{1}{2}$$

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 we are provided additional information 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.