CMPT 210: Probability and Computing

Lecture 18

Sharan Vaswani

March 19, 2024

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.

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 can be for any number of variables

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

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

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, $\mathrm{PDF}_{X,Y}[x,y] = \mathrm{PDF}_X[x]\,\mathrm{PDF}_Y[y]$.

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

All of these events are disjoint, so we can sum up the probabilities.

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 $\mathrm{Range}[X] = \{x_1, x_2, \dots x_n\}$, $\mathrm{Range}[Y] = \{y_1, y_2, \dots y_n\}$, then for $x \in \mathrm{Range}(X)$,
$[X = x] = [X = x \cap y = y_1] \cup [X = x \cap y = y_2] \cup \dots \cup [X = x \cap y = y_n]$ True for independent and dependent variables
$\implies \Pr[X = x] = \Pr[X = x \cap y = y_1] + \Pr[X = x \cap y = y_2] + \dots + \Pr[X = x \cap y = y_n]$.
A single distribution can be obtained by summing over the values that another distribution can take.
$\implies \mathrm{PDF}_X[x] = \sum_i \mathrm{PDF}_{X,Y}[x, y_i]$.
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 the batteries are distinct and we let X and Y denote, respectively, the number of new and used but still working batteries that are chosen, completely specify $\text{PDF}_{X,Y}$.

X: number of new batteries chosen
Y: number of used batteries chosen

We need the condition that X + Y <= 3

PDF(x, y) = 3cx * 4cy * 5c(3 - x - y) / 12c3
There are 12c3 ways to choose the three batteries

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 the batteries are distinct and we let X and Y denote, respectively, the number of new and used but still working batteries that are chosen, completely specify $\text{PDF}_{X,Y}$.

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

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

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

Should we have a term representing the defective batteries? I thought that if X + Y = 3, this term becomes 1.

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

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

For specifying a joint pdf, you must identify the range and identify the range.

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 $= P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

Pr(X = i) = 3ci * 9c(3 - i)/(12c3)

Since the 9 remaining batteries are distinct, we can find all the ways we can pick 3 - i batteries in the set of nine.

Questions?

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:

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

Expectation of a function.

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

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:

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

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 $(\mathbb{E}[X])^2$ We can let g(x) = x^2, and use the expectation of the function below.

E[X^2] = 1 * 1/6 + 4 * 1/6 + 9 * 1/6 + 16 * 1/6 + 25/6 + 36/6 = 91/6

(E(X))^2 = 441/36 = 49 /4

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 $(\mathbb{E}[X])^2$.

If asked if E[R1 * R2] = E[R1]E[R2] given that R1 and R2 are not independent,
let R1 = x and R2 = x.
Then, you will have that E[x^2] = (E[x])^2, which we know this is false.

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

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

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

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.

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:

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

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,

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

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

Variance

Standard way to measure the deviation from the mean is to calculate the *variance*. For r.v. X,

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

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 $\mathrm{Var}[X]$.

Var[X] = E[X^2] - E[X]^2 = p - p^2 = p(1 - p)

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

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

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).

If p is 1 or 0, then there is no variance in the results we obtain.

Variance

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

Variance

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

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

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

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

Variance

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

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

Variance

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

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

Variance

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

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

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

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

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

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

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

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

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

Back to throwing dice

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

Var[X] = E[X^2] - (E[X])^2 = 91/6 - 49/4 = (182 - 147) / 12 = (35/12)

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,

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

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

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

Back to throwing dice

Q: For a standard dice, if X is the r.v. equal to the number that comes up, compute 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 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:

$$W = 0 \qquad \text{(with } p = 1)$$

$$Y = -1 \qquad \text{(with } p = 1/2)$$

$$= +1 \qquad \text{(with } p = 1/2)$$

$$Z = -1000 \qquad \text{(with } p = 1/2)$$

$$= +1000 \qquad \text{(with } p = 1/2)$$

Var[W] = E[W^2] - (E[W])^2 - 0 - 0 = 0

Var[Y] = 1 - 0 = 1

] = 1,000,000 = (-1000)^2 * 1/2 + (1000)^2 * 1/2 - (E[Z])^2 = 1,000,000

Variance - Examples

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

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

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".

Variance - Examples

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

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

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 $\text{Var}[R]$.

E[R] = E[R^2] - (E[R])^2
= E[R^2] - (1/p)^2

Variance - Examples

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

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

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,

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

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,

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

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,

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

Variance - Examples

Continuing from the previous slide,

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

Putting everything together,

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

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

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

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

Variance - Examples

Continuing from the previous slide,

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

Putting everything together,

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

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

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

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