# Lecture 6

- **Read:** Chapter 4.6-4.10.
- Multiple Discrete RVs
  - Expectations of Functions of Two Random Variables
  - Covariance and Correlation
  - Conditioning
    - Conditioning a Joint PMF by an Event
    - Conditional PMF
  - Independent Random Variables
  - More Than Two Discrete Random Variables

# Expectation of Functions

- In many situations we need to know only the expected value of a derived random variable rather than the entire probability model.
- In these situations, we can obtain the expected value directly from the joint PMF of the random variable pair.
  - We do not have to compute the PMF of the derived random variable.

# Expectation of Functions

- **Theorem:** If $W = g(X, Y)$, then

$$E[W] = E[g(X, Y)] = \sum_x \sum_y g(x, y) p_{X,Y}(x, y)$$

- **Theorem:** If $g(X, Y) = g_1(X, Y) + g_1(X, Y) + ... + g_n(X, Y)$, then

$$E[g(X, Y)] = E[g_1(X, Y)] + ... + E[g_n(X, Y)]$$

"Expectation of the sum is equal to the sum of expectations!"

# Expectation of Functions: Proof of Theorem

$$
\begin{aligned}
E[g(X,Y)] &= \sum_x \sum_y g(x,y) p_{X,Y}(x,y) \\
&= \sum_x \sum_y (g_1(x,y) + \ldots + g_n(x,y)) p_{X,Y}(x,y) \\
&= \sum_x \sum_y [g_1(x,y) p_{X,Y}(x,y) + \ldots + g_n(x,y) p_{X,Y}(x,y)] \\
&= \underbrace{\sum_x \sum_y g_1(x,y) p_{X,Y}(x,y)}_{E[g_1(X,Y)]} + \ldots \\
&\qquad + \underbrace{\sum_x \sum_y g_n(x,y) p_{X,Y}(x,y)}_{E[g_n(X,Y)]} \\
&= E[g_1(X,Y)] + \ldots + E[g_n(X,Y)]
\end{aligned}
$$

# Expectation of Functions: Sum of Two RVs

- Let $(X, Y)$ have joint PMF $p_{X,Y}(x, y)$.
- Then,

$$E[X+Y] = E[X] + E[Y]$$

$$\begin{aligned} Var[X+Y] &= E[(X+Y-\mu_X-\mu_Y)^2] = E[(X-\mu_X+Y-\mu_Y)^2] \\ &= E[(X-\mu_X)^2 + 2(X-\mu_X)(Y-\mu_Y) + (Y-\mu_Y)^2] \\ &= Var[X] + Var[Y] + 2E[(X-\mu_X)(Y-\mu_Y)] \end{aligned}$$

# Covariance

- **Definition:(Covariance)** Covariance of two RVs $(X, Y)$ is

$$\begin{aligned} Cov[X, Y] &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY - X\mu_Y - \mu_X Y + \mu_X \mu_Y] \\ &= E[XY] - E[X\mu_Y] - E[\mu_X Y] + E[\mu_X \mu_Y] \end{aligned}$$

$$E[X\mu_Y] = \sum_x (x\mu_Y) p_X(x) = \mu_Y E[X] = \mu_X \mu_Y$$

$$\begin{aligned} Cov[X, Y] &= E[XY] - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y \\ &= E[XY] - \mu_X \mu_Y \end{aligned}$$

- **Note:** Suppose $Cov[X, Y] > 0$. This suggests that on average,
  - either $X > \mu_X$ and $Y > \mu_Y$
  - or $X < \mu_X$ and $Y < \mu_Y$

- **Interpretation:** If $Cov[X, Y] > 0$, then $X - \mu_X$ and $Y - \mu_Y$ tend to stray on the same side of their means. If $Cov[X, Y] < 0$, they tend to stray in opposite directions.

# Correlation

- **Recall:** $Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y]$
- **Definition:(Correlation)** Correlation of $X$ and $Y$ is $E[X \cdot Y]$.
- **Note:** If $E[X] = E[Y] = 0$, then $Cov[X,Y] = E[X \cdot Y]$.
- **Definition:(Orthogonal)** $X$ and $Y$ are said to be orthogonal if $E[X \cdot Y] = 0$.

# Correlation

- **Definition:(Uncorrelated RVs)** $(X, Y)$ are such that $Cov[X, Y] = 0$ or alternatively $E[XY] = E[X]E[Y]$.
- **Note:** $Cov[X, X] = Var[X]$
- **Recall:** $Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$
- **Note:** If $X$ and $Y$ are uncorrelated
$$Var[X + Y] = Var[X] + Var[Y]$$

# Correlation Coefficient

- **Definition:(Correlation Coefficient)** $\rho_{X,Y}$

$$\rho_{X,Y} = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$$

- **Theorem:** The correlation coefficient is normalized, i.e.,

$$-1 \leq \rho_{X,Y} \leq 1$$

(For proof, refer to text.)

- **Theorem:** If $Y = aX + b$, then

$$\rho_{X,Y} = \begin{cases} -1 & , a < 0 \\ 1 & , a > 0 \\ 0 & , a = 0 \end{cases}$$

# Correlation Coefficient: Proof of Theorem

$$Cov[X,Y] = E[(X-\mu_X)(Y-\mu_Y)] = E[(X-\mu_X)(aX+b-a\mu_X-b]$$

$$[\quad \mu_Y = E[Y] = E[aX+b] = aE[X]+b = a\mu_X + b \quad]$$

$$Cov[X,Y] = aE[(X-\mu_X)^2] = aVar[X]$$

$$Var[Y] = a^2 Var[X]$$

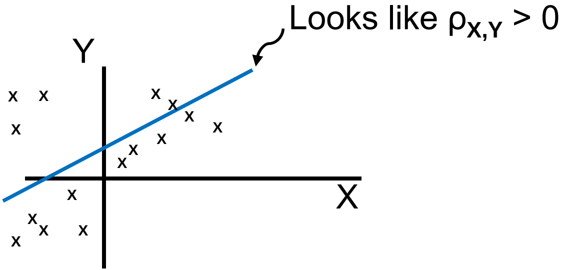
not $a$!

$$\rho_{X,Y} = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}} = \frac{aVar[X]}{\sqrt{a^2 Var[X]^2}}$$

$$= \frac{a}{\sqrt{a^2}} = \frac{a}{|a|} = \begin{cases} 1 & \text{, if } a > 0 \\ -1 & \text{, if } a < 0 \end{cases}$$

# Correlation Coefficient: Example

- Collect data $(x_1, y_1), \ldots, (x_n, y_n)$.




- And if $\rho_{X,Y} = 1$, the relationship between $X$ and $Y$ might be appropriately modeled by a straight line, with positive slope!

# Conditioning

- **Definition:(Conditioning Joint PMF by an Event)** The conditional joint PMF of $(X, Y)$ given some event $B$, a set on the $x$-$y$ plane, is defined as

$$
\begin{aligned}
p_{X,Y|B}(x,y) &= P[X = x, Y = y|B] \\
&= \frac{P[\{X = x\} \cap \{Y = y\} \cap B]}{P[B]} \\
&= \begin{cases} \frac{p_{X,Y}(x,y)}{P[B]} & , (x,y) \in B \\ 0 & , \text{otherwise} \end{cases}
\end{aligned}
$$

- **Definition:(Conditioning an RV Based on Another)** The conditional PMF of $X$ given $Y$ is

$$
\begin{aligned}
p_{X|Y}(x|y) &= P[X = x|Y = y] \\
&= \frac{P[\{X = x\} \cap \{Y = y\}]}{P[Y = y]} = \frac{p_{X,Y}(x,y)}{p_Y(y)}
\end{aligned}
$$

# Conditional Expectation

- **Definition:(Conditional Expectation of a Function)** Conditional expectation of $g(X, Y)$ given $B$

$$E[g(X, Y)|B] = \sum_x \sum_y g(x, y) \cdot p_{X,Y|B}(x, y)$$

- **Definition:(Conditional Expectation)** Conditional expectation of $X$ given $Y = y$

$$E[X|Y = y] = \sum_x x p_{X|Y}(x|y)$$

# Independent Random Variables

- **Definition:** $X$ and $Y$ are independent if $\{\mathrm{X}=\mathrm{x}\}$ and $\{\mathrm{Y}=\mathrm{y}\}$ are independent events for all $x$, $y$.

$$P[\{X=x\}\cap\{Y=y\}] = P[X=x]\cdot P[Y=y]$$
$$\Rightarrow p_{X,Y}(x,y) = p_X(x)\cdot p_Y(y)$$

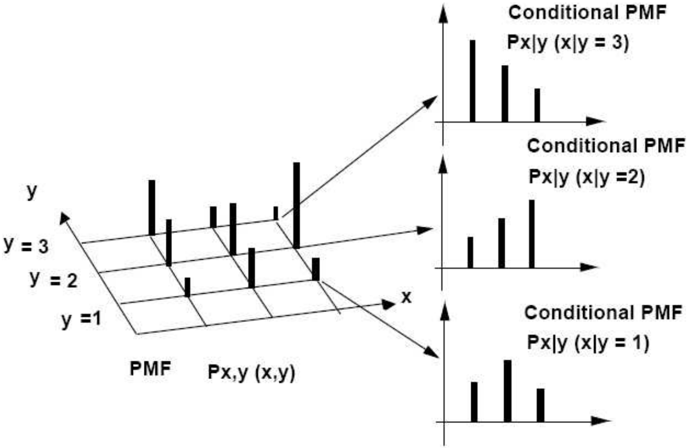
Equivalently, since for all $x$, $y$ such that $p_Y(y)>0$

$$p_{X,Y}(x,y) = p_{X|Y}(x|y)\cdot p_Y(y)$$
$$\underbrace{=}_{\text{independent}} \quad p_X(x)p_Y(y)$$

we have independence when $p_{X|Y}(x|y) = p_X(x)$.

  - **Interpretation:** knowledge of the experimental value of $Y$ does not affect PMF of $X$.
  - Given two functions $g$ and $h$, if $X$ and $Y$ are independent, so are $g(X)$ and $h(Y)$.
  - **Remark:** Independence iff can factorize the joint PMFs into functions of $x$ and $y$ alone.

# Visualizing Conditioning



# Summary

Joint PMF: $p_{X,Y}(x,y) = P[X = x, Y = y]$

if independent

marginal PMF

$$p_X(x) = \sum_{y \in S_Y} p_{X,Y}(x,y)$$

$$p_Y(y) = \sum_{x \in S_X} p_{X,Y}(x,y)$$

conditional PMF

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$$

# Joint Distribution from Marginals

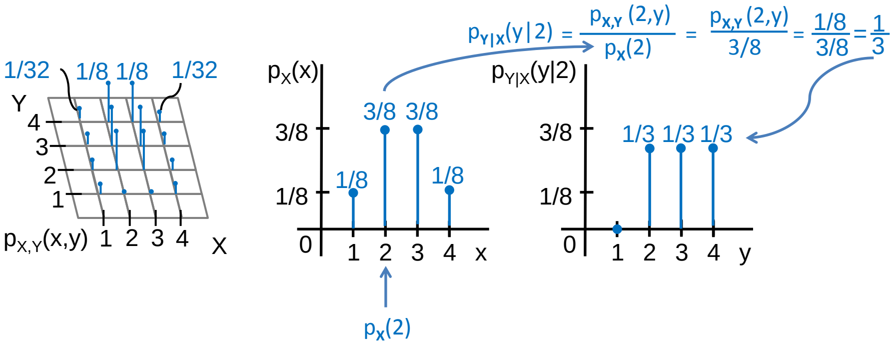
- **Question:** Given the marginal distributions of two RVs, can you reconstruct the joint distribution?
- **Answer:** Yes, if they are independent!

# Summary: Independent RVs

Independent RVs: $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$

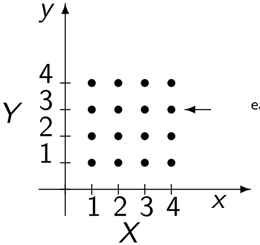
$$\text{Alternatively, } \begin{aligned} p_{X|Y}(x|y) &= p_X(x) \\ p_{Y|X}(y|x) &= p_Y(y) \end{aligned}$$

# Summary: Visualizing Conditioning




- If independent, then given the value of $X$, the distribution of $Y$ would not change.

# Independent RVs: Example

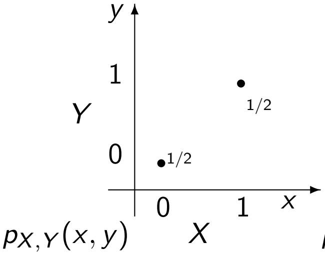



$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{1/16}{1/4} = p_X(x)$$

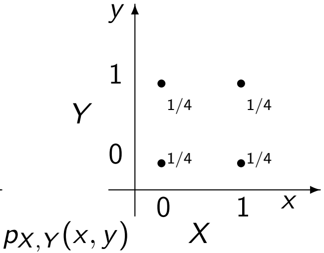
independent RVs

# Support (Range) of Independent RVs

- **Note:** Independent RVs will have a range (support) $S_{X,Y} = S_X \times S_Y$, i.e., rectangular.

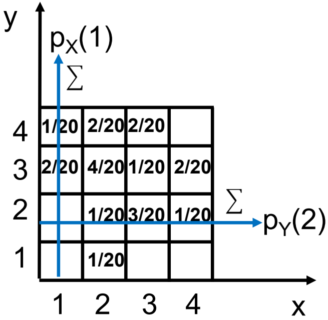



dependent




independent

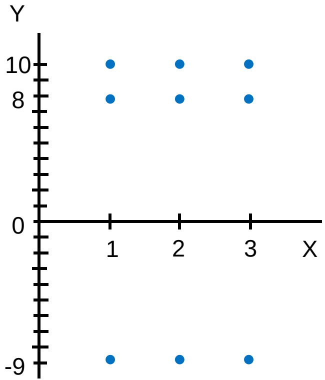
# Support of RVs




- Consider the joint PMF of the two RVs in the table above.
- Are they independent?
- Does a choice of $X = x$, e.g., 1 or 2, impact the possible values $Y$ can take?

# Do Only Independent RVs Have Rectangular Supports?

- **Example:**




$$X = \begin{cases} 1 & \text{, with probability } 1/6 \\ 2 & \text{, with probability } 2/6 \\ 3 & \text{, with probability } 3/6 \end{cases}$$

$$Y = \begin{cases} -9 & \text{, with probability } 1/3 \\ 8 & \text{, with probability } 1/3 \\ 10 & \text{, with probability } 1/3 \end{cases}$$

- **Warning:** Ranges that are "rectangular" in this sense may or may not correspond to independent random variables!

# Conditional Expected Values

- **Idea:** Given $Y = y$, the PMF of $X$ is $p_{X|Y}(x|y)$, so the average of $X$ is now given by

$$E[X|Y = y] = \sum_{x \in S_X} x p_{X|Y}(x|y)$$

We can generalize this definition and define the conditional expectation of a function of the random variable $X$...

# Conditional Expected Value of a Function

- **Definition:(Conditional Expected Value of a Function)**
  For any $y \in S_Y$, the conditional expected value of $g(X)$ given $Y = y$ is

$$E[g(X)|Y = y] = \sum_{x \in S_X} g(x) p_{X|Y}(x|y)$$

Special Cases:

1. E[X|Y= y]
2. Conditional Variance

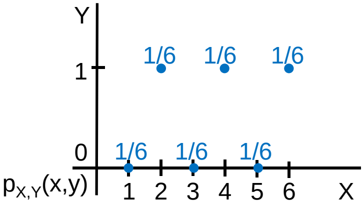
$$\begin{aligned} Var[X|Y = y] &= E[(X - E[X|Y = y])^2 | Y = y] \\ &= \sum_{x \in S_X} (x - E[X|Y = y])^2 p_{X|Y}(x|y) \end{aligned}$$

# Conditional Expectation: Example (I)

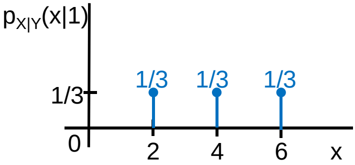
- Let $X =$ outcome of roll of die

$$Y = \begin{cases} 1 & \text{, if } X \text{ was even} \\ 0 & \text{, if } X \text{ was odd} \end{cases}$$

- What are $p_{X,Y}(x,y)$, $p_{X|Y}(x|1)$?




$$p_{X|Y}(x|1) = \frac{p_{X,Y}(x,1)}{p_Y(1)} = \frac{p_{X,Y}(x,1)}{1/2}$$



# Conditional Expectation: Example (II)

- Let $X =$ outcome of roll of die

$$Y = \begin{cases} 1 & \text{, if } X \text{ was even} \\ 0 & \text{, if } X \text{ was odd} \end{cases}$$

- What are $E[X|Y = 1]$ and $Var[X]$?

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

$$E[X|Y = 1] = \sum_{x \in S_X} x p_{X|Y}(x|1) = 2 \cdot \frac{1}{3} + 4 \cdot \frac{1}{3} + 6 \cdot \frac{1}{3} = 4$$

$$\begin{aligned} Var[X] &= E[(X - E[X])^2] = E[(X - 3.5)^2] = E[X^2] - (E[X])^2 \\ &= \frac{1}{6}[(1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + (4 - 3.5)^2 \\ &\qquad + (5 - 3.5)^2 + (6 - 3.5)^2] \end{aligned}$$

# Conditional Expectation: Example (III)

- Let $X =$ outcome of roll of die

$$Y = \begin{cases} 1 & \text{, if } X \text{ was even} \\ 0 & \text{, if } X \text{ was odd} \end{cases}$$

- What is $Var[X|Y = 1]$?

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

$$\begin{aligned} Var[X|Y = y] &= E[(X - E[X|Y = y])^2|Y = y] \\ &= \sum_{x \in S_X} (x - E[X]|Y = y)^2 p_{X|Y}(x|y) \end{aligned}$$

$$\begin{aligned} Var[X|Y = 1] &= E[X^2|Y = 1] - (E[X|Y = 1])^2 \\ &= 4 \cdot \frac{1}{3} + 16 \cdot \frac{1}{3} + 36 \cdot \frac{1}{3} - 16 \end{aligned}$$

# Conditional Expectation

- **Definition:(Conditional Expectation)** Let $g(y) = E[X|Y = y]$. Then, $g(Y)$ is called the conditional expectation of $X$ given $Y$ and is written as $E[X|Y]$.

- **Warning:** $E[X|Y] = g(Y)$ is a function of an RV, $Y$, so it is also an RV, i.e., necessarily just a number.

  Since $E[X|Y]$ is an RV, we should be able to take its average!

- **Theorem:**(Iterated Expectation) $E[E[X|Y]] = E[X]$

  - The significance of the iterated expectation is that there are problems in which calculating $E[X]$ directly is much more difficult than first calculating $E[X|Y]$ and then using the iterated expectation to find $E[X]$.

# Iterated Expectation: Proof of Theorem

- **Theorem:** E[E[X|Y]] = E[X]
- **Proof:**

$$
\begin{aligned}
E[E[X|Y]] &= E[g(Y)] \\
&= \sum_y g(y) p_Y(y) \\
&= \sum_y E[X|Y=y] p_Y(y) \\
&= \sum_y \left( \sum_x x p_{X|Y}(x|y) \right) p_Y(y) \\
\sum_y \sum_x x p_{X|Y}(x|y) p_Y(y) &= \sum_y \sum_x x p_{X,Y}(x,y) \\
&= \sum_x x \sum_y p_{X,Y}(x,y) \\
&= \sum_x x p_X(x) = E[X]
\end{aligned}
$$

# Iterated Expectation: Example

- Suppose
  $Y = 0 \Rightarrow$ Male, P[Y$=$ 0]$=$ 1/2
  $Y = 1 \Rightarrow$ Female

- Let $X =$ age of person
  $E[X|Y = 0] =$ average age of males
  $Var[X|Y = 1] =$ variance of ages of females

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

$$E[X|Y] = g(Y) = \begin{cases} E[X|Y=0] & \text{, with probability } 1/2 \\ E[X|Y=1] & \text{, with probability } 1/2 \end{cases}$$

$$\begin{aligned} g(Y) &= E[X|Y=y] \\ E[E[X|Y]] &= E[X|Y=0]P[Y=0] + E[X|Y=1]P[Y=1] \\ &= E[X] \end{aligned}$$

# Facts About Independent RVs

If $X$ and $Y$ are independent

- $E[XY] = E[X]E[Y]$, so independent $\Longrightarrow$ uncorrelated, i.e., $Cov[X, Y] = 0$.
- $E[X|Y = y] = E[X]$ for all $y \in S_Y$
- $E[Y|X = x] = E[Y]$ for all $x \in S_X$
- $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$
- $Var[X + Y] = Var[X] + Var[Y]$

# Independence of Several Random Variables

- $X$, $Y$, and $Z$ are independent if
$$p_{X,Y,Z}(x,y,z) = p_X(x)p_Y(y)p_Z(z)$$
for all $x$, $y$, and $z$.
- If so, $f(X)$, $g(Y)$, and $h(Z)$ would also be independent.
- Typically, $f(X,Y)$ and $g(Y,Z)$ would not.
- Also, $Var[X+Y+Z] = Var[X] + Var[Y] + Var[Z]$.
- **Example:** Suppose $X_i \sim$ Bernoulli($p$), and let $Y = \sum_{i=1} X_i$. Find its variance.

# More Than Two Discrete Random Variables: Joint PMF

- **Definition:(Joint PMF of $N$ Random Variables)** The joint PMF of the discrete random variables $X_1, \ldots, X_n$ is

$$p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P[X_1 = x_1, \ldots, X_n = x_n]$$

# More Than Two Discrete Random Variables: Probability of an Event $A$

- **Theorem:** The probability of an event $A$ expressed in terms of the discrete random variables $X_1, ..., X_n$ is

$$P[A] = \sum_{(x_1,...,x_n)\in A} p_{X_1,...,X_n}(x_1, ..., x_n)$$

# More Than Two Discrete Random Variables: Marginal PMFs

- **Theorem:** For a joint PMF of four random variables, $p_{W,X,Y,Z}(w,x,y,z)$, some marginal PMFs are

$$
\begin{aligned}
p_{X,Y,Z}(x,y,z) &= \sum_{w \in S_W} p_{W,X,Y,Z}(w,x,y,z) \\
p_{W,Z}(w,z) &= \sum_{x \in S_X} \sum_{y \in S_Y} p_{W,X,Y,Z}(w,x,y,z) \\
p_{Y,Z}(w,z) &= \sum_{w \in S_W} \sum_{x \in S_X} p_{W,X,Y,Z}(w,x,y,z) \\
p_X(x) &= \sum_{w \in S_W} \sum_{y \in S_Y} \sum_{z \in S_Z} p_{W,X,Y,Z}(w,x,y,z)
\end{aligned}
$$

# $N$ Independent Random Variables

- **Definition:($N$ Independent Random Variables)** The discrete random variables $X_1, ..., X_n$ are independent if and only if

$$p_{X_1,...,X_n}(x_1, ..., x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$$

for all $x_1, ..., x_n$.

# Expectations

- **Theorem:** The expected value of $g(X_1, ..., X_n)$ satisfies

$$E[g(X_1, ..., X_n)] = \sum_{x_1 \in S_{X_1}} ... \sum_{x_n \in S_{X_n}} g(x_1, ..., x_n) p_{X_1,...,X_n}(x_1, ..., x_n)$$

- **Theorem:** When $X_1, ..., X_n$ are independent discrete random variables,

$$E[g(X_1)g(X_2) \cdots g(X_n)] = E[g(X_1)]E[g(X_2)] \cdots E[g(X_n)]$$