EE 325: Probability and Random Processes
Module 3: Expectations, Functions of a Random Variable, Higher Moments

D Manjunath

Networking Lab
Department of Electrical Engineering
Indian Institute of Technology

September 28, 2020

Topics in Module 3

- Expectation of random variable.
- Functions of a random variable of the form $g(\mathsf{X})$.
- Moments of a random variable.
- Moment generating functions and characteristic functions of random variables.
- Mostly from Chapter 5 of the text; Inequalities will be in the next module.

Expectation

- In a coin toss, let $H = 1$ and $T = -1$ be the values associated with heads and tails, i.e., the random variable $\mathsf{X} \in \{-1, 1\}$
- Toss a coin (not necessarily a fair one, call it the Shakuni coin) N times; Let N_h be the number of heads in the N tosses.
- Sample realisation: H T T T H H H T H H T H H T H T
- The sample mean is

$$\frac{N_H - N_T}{N} = \frac{N_H(+1) + N_T(-1)}{N}$$

$$= \frac{N_H}{N}(+1) + \frac{N_T}{N}(-1)$$

- As $N \to \infty$, $\frac{N_H}{N} \to p$ and $\frac{N_T}{N} \to (1-p)$; average, denoted by $\bar{\mathsf{X}}$, will be

$$p_{+1}(+1) + p_{-1}(-1)$$

- Generalising, for N samples of a discrete random variable, the sample mean would be

$$\bar{X} = \sum_i \frac{i\, N_i}{N} = \sum_i i \left(\frac{N_i}{N}\right)$$

Expectation

- In a coin toss, let $H = 1$ and $T = -1$ be the values associated with heads and tails, i.e., the random variable $\mathsf{X} \in \{-1, 1\}$
- Toss a coin (not necessarily a fair one, call it the Shakuni coin) N times; Let N_h be the number of heads in the N tosses.
- Sample realisation: H T T T H H H T H H T H H T H T
- The sample mean is

$$\begin{aligned} \frac{N_H - N_T}{N} &= \frac{N_H(+1) + N_T(-1)}{N} \\ &= \frac{N_H}{N}(+1) + \frac{N_T}{N}(-1) \end{aligned}$$

- As $N \to \infty$, $\frac{N_H}{N} \to p$ and $\frac{N_T}{N} \to (1-p)$; average, denoted by $\bar{\mathsf{X}}$, will be

$$p_{+1}(+1) + p_{-1}(-1)$$

- Generalising, for N samples of a discrete random variable, the sample mean would be

$$\bar{X} = \sum_i \frac{i\, N_i}{N} = \sum_i i \left(\frac{N_i}{N}\right)$$

Expectation

- Formally, the expectation of a discrete random variable is defined as

$$\mathsf{E}(\mathsf{X}) = \sum_i i \, p_{\mathsf{X}}(i)$$

- For continuous random variables

$$\mathsf{E}(\mathsf{X}) = \int_{-\infty}^{\infty} x \, f_{\mathsf{X}}(x) \, dx$$

- It is possible that the summation, or the integration, is not finite. In this case

Example Expectations

- In a coin toss, let $H = 1$ and $T = -1$ be the values associated with heads and tails, i.e., the random variable $\mathsf{X} \in \{-1, 1\}$

$$\mathsf{Pr}(\mathsf{X} = 1) = 0.3 \quad \mathsf{Pr}(\mathsf{X} = -1) = 0.7$$
$$\mathsf{E}(X) = 0.3 - 0.7 = -0.4$$

- $0 - 1$ Bernouilli Random Variable:

$$\mathsf{X} \in \{0, 1\}$$
$$p_{\mathsf{X}}(1) = \alpha$$
$$p_{\mathsf{X}}(0) = 1 - \alpha$$
$$\mathsf{E}(X) = \alpha$$

Example Expectations

- In a coin toss, let $H = 1$ and $T = -1$ be the values associated with heads and tails, i.e., the random variable $\mathsf{X} \in \{-1, 1\}$

$$\begin{aligned} \mathsf{Pr}(\mathsf{X} = 1) &= 0.3 \quad \mathsf{Pr}(\mathsf{X} = -1) = 0.7 \\ \mathsf{E}(X) &= 0.3 - 0.7 = -0.4 \end{aligned}$$

- $0 - 1$ Bernouilli Random Variable:

$$\begin{aligned} &\mathsf{X} \in \{0, 1\} \\ p_{\mathsf{X}}(1) &= \alpha \\ p_{\mathsf{X}}(0) &= 1 - \alpha \\ &\mathsf{E}(X) = \alpha \end{aligned}$$

Example Expectations

- In a coin toss, let $H = 1$ and $T = -1$ be the values associated with heads and tails, i.e., the random variable $\mathsf{X} \in \{-1, 1\}$

$$\begin{aligned} \mathsf{Pr}(\mathsf{X} = 1) &= 0.3 \quad \mathsf{Pr}(\mathsf{X} = -1) = 0.7 \\ \mathsf{E}(X) &= 0.3 - 0.7 = -0.4 \end{aligned}$$

- $0 - 1$ Bernouilli Random Variable:

$$\begin{aligned} &\mathsf{X} \in \{0, 1\} \\ &p_{\mathsf{X}}(1) = \alpha \\ &p_{\mathsf{X}}(0) = 1 - \alpha \\ &\mathsf{E}(X) = \alpha \end{aligned}$$

Example Expectations

- In a coin toss, let $H = 1$ and $T = -1$ be the values associated with heads and tails, i.e., the random variable $\mathsf{X} \in \{-1, 1\}$

$$
\begin{aligned}
\mathsf{Pr}(\mathsf{X} = 1) &= 0.3 \quad \mathsf{Pr}(\mathsf{X} = -1) = 0.7 \\
\mathsf{E}(X) &= 0.3 - 0.7 = -0.4
\end{aligned}
$$

- $0 - 1$ Bernouilli Random Variable:

$$
\begin{aligned}
&\mathsf{X} \in \{0, 1\} \\
p_{\mathsf{X}}(1) &= \alpha \\
p_{\mathsf{X}}(0) &= 1 - \alpha \\
&\mathsf{E}(X) = \alpha
\end{aligned}
$$

Example Expectations

- In a coin toss, let $H = 1$ and $T = -1$ be the values associated with heads and tails, i.e., the random variable $\mathsf{X} \in \{-1, 1\}$

$$\mathsf{Pr}(\mathsf{X} = 1) \;=\; 0.3 \quad \mathsf{Pr}(\mathsf{X} = -1) \;=\; 0.7$$
$$\mathsf{E}(X) \;=\; 0.3 - 0.7 \;=\; -0.4$$

- $0 - 1$ Bernouilli Random Variable:

$$\mathsf{X} \in \{0, 1\}$$
$$p_{\mathsf{X}}(1) \;=\; \alpha$$
$$p_{\mathsf{X}}(0) \;=\; 1 - \alpha$$
$$\mathsf{E}(X) = \alpha$$

Example Expectations

- Two fair, six-sided dice are thrown; X is the sum of the values on the top face.

$p_X \downarrow \quad \mathsf{X} \rightarrow$	2	3	4	5	6	7
	1/36	2/36	3/36	4/36	5/36	6/36

$p_X \downarrow \quad \mathsf{X} \rightarrow$	8	9	10	11	12
	5/36	4/36	3/36	2/36	1/36

$$\begin{aligned} \mathsf{E}(X) &= (2+6+12+20+30+42+40+36+30+22+12)/36 \\ &= 252/36 \end{aligned}$$

- Binomial Random Variable: Count the number of Heads from N independent coin tosses.

$$\mathsf{X} \in \{0, 1, \ldots, N\}$$

$$\mathsf{Pr}(\mathsf{X} = k) = p_{\mathsf{X}}(k) = \begin{cases} \binom{N}{k} \alpha^k (1-\alpha)^{N-k} & \text{for } 0 \leq k \leq N \\ 0 & \text{otherwise} \end{cases}$$

Expectation of a Binomial RV

$$
\begin{aligned}
\mathsf{E}(X) &= \sum_{k=0}^{N} k \binom{N}{k} \alpha^k (1-\alpha)^{N-k} \\
&= \sum_{k=1}^{N} k \, \frac{N!}{k!\,(N-k)!} \alpha^k (1-\alpha)^{N-k} \\
&= N\alpha \sum_{k=1}^{N} \frac{(N-1)!}{(k-1)!\,(N-k)!} \alpha^{k-1} (1-\alpha)^{N-k} \\
&= N\alpha \sum_{k=1}^{N} \frac{(N-1)!}{(k-1)!\,(N-1)-(k-1)!} \alpha^{k-1} (1-\alpha)^{(N-1)-(k-1)} \\
&= N\alpha \sum_{k_1=0}^{N-1} \frac{(N-1)!}{k_1!\,((N-1)-k_1)!} \alpha^{k_1} (1-\alpha)^{(N-1)-k_1} \\
&= N\alpha \left[\sum_{k_1=0}^{N-1} \binom{N-1}{k_1} \alpha^{k_1} (1-\alpha)^{(N-1)-k_1} \right] = N\alpha
\end{aligned}
$$

Expectation of a Geometric RV

$$\mathsf{X} \in \{0, 1, \ldots\}$$

$$\Pr(\mathsf{X} = k) = p_{\mathsf{X}}(k) = \begin{cases} (1-\alpha)^k \alpha & \text{for } k \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathsf{E}(X) = \sum_{k=0}^{\infty} k(1-\alpha)^k \alpha$$

An identity

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$$

$$\sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$

$$\mathsf{E}(\mathsf{X}) = \alpha(1-\alpha)\sum_{k=0}^{\infty} k(1-\alpha)^{k-1} = \alpha(1-\alpha)\frac{1}{\alpha^2} = \frac{1-\alpha}{\alpha}$$

Expectation of a Geometric RV

$$\mathsf{X} \in \{0, 1, \ldots\}$$

$$\Pr(\mathsf{X} = k) = p_{\mathsf{X}}(k) = \begin{cases} (1-\alpha)^k \alpha & \text{for } k \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathsf{E}(X) = \sum_{k=0}^{\infty} k(1-\alpha)^k \alpha$$

An identity

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$$

$$\sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$

$$\mathsf{E}(\mathsf{X}) = \alpha(1-\alpha)\sum_{k=0}^{\infty} k(1-\alpha)^{k-1} = \alpha(1-\alpha)\frac{1}{\alpha^2} = \frac{1-\alpha}{\alpha}$$

Expectation of a Geometric RV

$$\mathsf{X} \in \{0, 1, \ldots\}$$

$$\Pr(\mathsf{X} = k) = p_{\mathsf{X}}(k) = \begin{cases} (1-\alpha)^k \alpha & \text{for } k \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathsf{E}(X) = \sum_{k=0}^{\infty} k(1-\alpha)^k \alpha$$

An identity

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$$

$$\sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$

$$\mathsf{E}(\mathsf{X}) = \alpha(1-\alpha)\sum_{k=0}^{\infty} k(1-\alpha)^{k-1} = \alpha(1-\alpha)\frac{1}{\alpha^2} = \frac{1-\alpha}{\alpha}$$

Expectation of a Poisson Random Variable

$$\mathsf{X} \in \{0, 1, \ldots\}$$

$$\Pr(\mathsf{X} = k) \;=\; p_{\mathsf{X}}(k) \;=\; \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda} & \text{for } k \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \mathsf{E}(\mathsf{X}) &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!}\, e^{-\lambda} \\ &= \lambda\, e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\ &= \lambda\, e^{-\lambda} \sum_{k_1=0}^{\infty} \frac{\lambda^{k_1}}{(k_1)!} \\ &= \lambda \end{aligned}$$

Expectation of uniform distribution

$$
\begin{aligned}
f_X(x) &= \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \\
\mathsf{E}(\mathsf{X}) &= \int_a^b x \frac{1}{b-a} dx \\
&= \frac{1}{b-a} \int_a^b x \, dx \\
&= \frac{1}{b-a} \frac{b^2 - a^2}{2} \\
&= \frac{a+b}{2}
\end{aligned}
$$

Expectation of exponential distribution

$$f_{\mathsf{X}}(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

To obtain the expectation from the definition, perform integration by parts with $u = x$ and $dv = \lambda e^{-\lambda x} dx$

$$\begin{aligned} \mathsf{E}(\mathsf{X}) &= \int_0^\infty x \lambda e^{-\lambda x} dx \\ &= x e^{-\lambda x} \Big|_0^\infty - \int_0^\infty -e^{-\lambda x} dx \\ &= x e^{-\lambda x} \Big|_0^\infty + \left(-\frac{1}{\lambda}\right) e^{-\lambda x} \Big|_0^\infty \\ &= \frac{1}{\lambda} \end{aligned}$$

Expectation of the Gaussian distribution

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Let $y = x - \mu$ and we get

$$\begin{aligned}
\mathsf{E(X)} &= \int_{-\infty}^{\infty} x \, \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (y+\mu) e^{-\frac{y^2}{2\sigma^2}} \, dy \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} y e^{-\frac{y^2}{2\sigma^2}} \, dy \; + \; \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \mu e^{-\frac{y^2}{2\sigma^2}} \, dy \\
&= 0 \; + \; \mu
\end{aligned}$$

Conditional expectation

- Conditional pdf and pmf can be used to define conditional expectation.

$$
\begin{aligned}
\mathsf{E}(\mathsf{X}|A) &= \int_{-\infty}^{\infty} x\, f_{\mathsf{X}|A}(x|A)\, dx \\
\mathsf{E}(\mathsf{X}|A) &= \sum_{x=-\infty}^{\infty} x\, p_{\mathsf{X}|A}(x|A)
\end{aligned}
$$

Function of a Random Variable

- Recall the notion of a function: Map the elements of the *domain* to the elements of the *range* (or *co-domain*).
- Consider a function $g(\cdot) : \Re \rightarrow \Re$.
- Let X be a random variable and let X be the variable for $g(\cdot)$, i.e., consider $\mathsf{Y} = g(\mathsf{X})$.
- Can view $g(\mathsf{X})$ as a map from Ω to the set of numbers and hence we can say $\mathsf{Y} = g(\mathsf{X})$ is a random variable.

Function of a Random Variable

- Consider $\Pr(\mathsf{Y} \in A)$; this is the same as $\Pr(g(\mathsf{X}) \in A)$.
- Let B be a subset of Ω defined as $B := \{\omega : g(\mathsf{X}(\omega)) \in A\}$.
- Clearly, $\Pr(\mathsf{Y} \in A) = \Pr(g(\mathsf{X}) \in A) = \Pr(B)$.
- The event $\{\mathsf{Y} \leq y\} = \{g(\mathsf{X}) \leq y\}q = \{\mathsf{X} \in g^{-1}(-\infty, y)\}$.
- This essentially implies

$$F_{\mathsf{Y}}(y) = \Pr(\mathsf{Y} \leq y) = \Pr\big(\mathsf{X} \in g^{-1}(-\infty, y)\big) .$$

Function of a Random Variable

- $\mathsf{Y} = g(\mathsf{X})$, g is not necessarily one-to-one.
- $f_{\mathsf{Y}}(y) \approx \Pr(y < \mathsf{Y} \leq y + dy)$.
- Need to obtain $f_{\mathsf{Y}}(y)$ from $f_{\mathsf{X}}(x)$.
- Assume $y = g(x)$ has k solutions: $x_1, \ldots x_k$.
- Event $\{y < \mathsf{Y} \leq y + dy\}$ can occur if any the following occur: $\{x_1 < \mathsf{X} \leq x_1 + dx_1\}, \cdots, \{x_k < \mathsf{X} \leq x_k + dx_k\}$
- These are mutually exclusive events. Thus

$$\Pr(y < \mathsf{Y} \leq y + dy) = \Pr(x_1 < \mathsf{X} \leq x_1 + dx_1) + \cdots + \Pr(x_k < \mathsf{X} \leq x_k + dx_k)$$

- Each of these events can be written as

$$\Pr(x_i < \mathsf{X} \leq x_i + dx_i) \quad = \quad f_{\mathsf{X}}(x_i)dx_i \; = f_{\mathsf{X}}(x_i)\frac{dy}{|g'(x_i)|}$$

- Thus

$$f_{\mathsf{Y}}(y)dy \quad = \quad \left(\sum_{i=1}^{k} \frac{f_{\mathsf{X}}(x_i)}{|g'(x_i)|}\right) dy \implies f_{\mathsf{Y}}(y) \; = \; \left(\sum_{i=1}^{k} \frac{f_{\mathsf{X}}(x_i)}{|g'(x_i)|}\right)$$

Problems

- $Y = X^2$.
- Specialise this for the case when X is a unit normal random variable.
- $Y = |X|$
- X is uniform in $[-5, 5]$ and

$$Y = \begin{cases} -1 & \text{if } X < -1 \\ X & \text{if } -1 \leq X \leq 1 \\ 1 & \text{if } X > 1 \end{cases}$$

- X is exponential with parameter λ and $Y = \lfloor X \rfloor$.
- X is uniformly distributed in $(0, \pi/2)$. $Y = \sin X$.
- X is uniformly distributed in $(0, 1)$. $Y = \sin^{-1} X$.

Expectation of a Function of a Random Variable

$$
\begin{aligned}
\mathsf{E}(g(\mathsf{X})) &= \sum_k g(k)\, p_{\mathsf{X}}(k) \\
\mathsf{E}(g(\mathsf{X})) &= \int g(x)\, f_{\mathsf{X}}(x)\, dx
\end{aligned}
$$

Analogously

$$
\begin{aligned}
\mathsf{E}(g(\mathsf{X},\mathsf{Y})) &= \sum_j \sum_k g(j,k)\, p_{\mathsf{X},\mathsf{Y}}(j,k) \\
\mathsf{E}(g(\mathsf{X},\mathsf{Y})) &= \int\int g(x,y)\, f_{\mathsf{X},\mathsf{Y}}(x,y)\, dx\, dy
\end{aligned}
$$

Expectation of a Function of a Random Variable

$$
\begin{aligned}
\mathsf{E}(g(\mathsf{X})) &= \sum_k g(k)\, p_{\mathsf{X}}(k) \\
\mathsf{E}(g(\mathsf{X})) &= \int g(x)\, f_{\mathsf{X}}(x)\, dx
\end{aligned}
$$

Analogously

$$
\begin{aligned}
\mathsf{E}(g(\mathsf{X},\mathsf{Y})) &= \sum_j \sum_k g(j,k)\, p_{\mathsf{X},\mathsf{Y}}(j,k) \\
\mathsf{E}(g(\mathsf{X},\mathsf{Y})) &= \int\int g(x,y)\, f_{\mathsf{X},\mathsf{Y}}(x,y)\, dx\, dy
\end{aligned}
$$

More about Expectations

- Non negative random variables have non negative expectations

$$\mathsf{X} \geq 0 \;\text{ i.e., } f_{\mathsf{X}}(x) = 0 \text{ for } x \leq 0, \quad \Longrightarrow \quad \mathsf{E}(\mathsf{X}) \geq 0$$

- Bounded random variables have bounded expectations

$$a \leq \mathsf{X} \leq b \;\text{ i.e., } f_{\mathsf{X}}(x) = 0 \text{ for } x \notin [a, b], \quad \Longrightarrow \quad a \leq \mathsf{E}(\mathsf{X}) \leq b.$$

- Expectation of a constant is a constant, i.e., if $\mathsf{X} = c$, then
- Expectation is a linear operation, i.e.

$$\begin{aligned}
\mathsf{E}(a\mathsf{X} + b) &= \int_{-\infty}^{\infty} (ax + b) f_{\mathsf{X}}(x) \\
&= a \int_{-\infty}^{\infty} x\, f_{\mathsf{X}}(x)\, dx + b \int_{-\infty}^{\infty} f_{\mathsf{X}}(x)\, dx \\
&= a\bar{\mathsf{X}} + b
\end{aligned}$$

- What is $\mathsf{E}\left(\mathsf{X} - \bar{\mathsf{X}}\right)$?

More about Expectations

- Compare $\mathsf{E}(\mathsf{X}-\bar{\mathsf{X}})$ with $\mathsf{E}(\mathsf{X}-a)$ for a constant a.

$$
\begin{aligned}
\mathsf{E}\left((\mathsf{X}-a)^2\right) &= \mathsf{E}\left(((\mathsf{X}-\bar{\mathsf{X}})+(\bar{\mathsf{X}}-a))^2\right)\\
&= \mathsf{E}\left((\mathsf{X}-\bar{\mathsf{X}})^2 + (\bar{\mathsf{X}}-a)^2 + 2\,(\mathsf{X}-\bar{\mathsf{X}})\,(\bar{\mathsf{X}}-a)\right)\\
&= \mathsf{E}\left((\mathsf{X}-\bar{\mathsf{X}})^2\right) + \mathsf{E}\left((\bar{\mathsf{X}}-a)^2\right) + 2\mathsf{E}((\mathsf{X}-\bar{\mathsf{X}})\,(\bar{\mathsf{X}}-a)\\
&= \mathsf{E}\left((\mathsf{X}-\bar{\mathsf{X}})^2\right) + (\bar{\mathsf{X}}-a)^2 + 2\,(\bar{\mathsf{X}}-a)\,\mathsf{E}((\mathsf{X}-\bar{\mathsf{X}}))\\
&= \mathsf{E}\left((\mathsf{X}-\bar{\mathsf{X}})^2\right) + (\bar{\mathsf{X}}-a)^2\\
\mathsf{E}\left((\mathsf{X}-a)^2\right) &\geq \mathsf{E}\left((\mathsf{X}-\bar{\mathsf{X}})^2\right)
\end{aligned}
$$

More Expectations: Moments

- k-th moment:

$$\mathsf{E}(\mathsf{X}^k) = \sum_i i^k \, p_{\mathsf{X}}(i)$$

$$\mathsf{E}(\mathsf{X}^k) = \int_{-\infty}^{\infty} x^k \, f_{\mathsf{X}}(x) \, dx$$

- k-th central moment: $(\mathsf{X} - \bar{\mathsf{X}})^k$

$$\mathsf{E}((\mathsf{X} - \bar{\mathsf{X}})^k) = \sum_i (i - \bar{\mathsf{X}})^k \, p_{\mathsf{bX}}(i)$$

$$\mathsf{E}(\mathsf{X}^k) = \int_{-\infty}^{\infty} (x - \bar{\mathsf{X}})^k \, f_{\mathsf{X}}(x) \, dx$$

Variance of a Random Variable

- Variance of a random variable X is defined as the expectation of $(\mathsf{X} - \mathsf{E}(X))^2$ and we will denote it by $\mathsf{VAR}(\mathsf{X})$.

$$
\begin{aligned}
\mathsf{VAR}(\mathsf{X}) &= \mathsf{E}\left((\mathsf{X} - \bar{\mathsf{X}})^2\right) \\
&= \mathsf{E}(\mathsf{X}^2 - 2\mathsf{X}\bar{\mathsf{X}} + \bar{\mathsf{X}}^2) \\
&= \mathsf{E}(\mathsf{X}^2) - 2\bar{\mathsf{X}}\mathsf{E}(\mathsf{X}) + \bar{\mathsf{X}}^2 \\
&= \mathsf{E}(\mathsf{X}^2) - \bar{\mathsf{X}}^2
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{E}(\mathsf{X}^2) &= \sum_k k^2 p_{\mathsf{X}}(k) \\
&= \int x^2 f_{\mathsf{X}}(x)\, dx
\end{aligned}
$$

Variance: Examples

- In a coin toss, let $H = 1$ and $T = -1$ be the values associated with heads and tails, i.e., the random variable $\mathsf{X} \in \{-1, 1\}$
  $\mathsf{Pr}(\mathsf{X} = 1) = 0.3$ and $\mathsf{Pr}(\mathsf{X} = -1) = 0.7$
- Toss two indpendent coins; Along the lines of the previous random variable, let $HH = 2$, $HT = TH = 0$ and $TT = -2$.
  $\mathsf{Pr}(\mathsf{X} = 2) = 0.09$, $\mathsf{Pr}(\mathsf{X} = 0) = 0.42$, and $\mathsf{Pr}(\mathsf{X} = -1) = 0.49$

Variance of a Random Variable

- What is the variance of a constant?
- Variance of a scalar multiple of a random variable

$$
\begin{aligned}
\mathsf{VAR}(a\mathsf{X}) &= \mathsf{E}\left((a\mathsf{X} - \mathsf{E}(a\mathsf{X}))^2\right) \\
&= \mathsf{E}\left((a\mathsf{X} - a\mathsf{E}(\mathsf{X}))^2\right) \\
&= \mathsf{E}\left(a^2\,(\mathsf{X} - \mathsf{E}(\mathsf{X}))^2\right) \\
&= a^2\mathsf{E}\left((\mathsf{X} - \mathsf{E}(\mathsf{X}))^2\right) \\
&= a^2\mathsf{VAR}(\mathsf{X})
\end{aligned}
$$

- $0 - 1$ Bernouilli Random Variable:

$$
\begin{aligned}
&\mathsf{X} \in \{0, 1\} \\
&p_{\mathsf{X}}(1) = \alpha, \;\; p_{\mathsf{X}}(0) = 1 - \alpha \\
&\mathsf{E}(\mathsf{X}) = \alpha, \;\; \mathsf{E}(\mathsf{X}^2) = \alpha \\
&\mathsf{VAR}(\mathsf{X}) = \alpha - \alpha^2 = \alpha(1 - \alpha)
\end{aligned}
$$

Variance of a Random Variable

- Geometric(α) random variable: $\frac{1-\alpha}{\alpha^2}$
- Binomial(N, α) random variable: $N\alpha(1-\alpha)$
- Poisson(λ) random variable: λ
- Uniform($0, 1$) random variable: $1/12$
- Exponential(λ) random variable: $1/\lambda^2$
- Gaussian(μ, σ) random variable: σ

More Expectations: Generating and Characteristic Functions

- If X is a discrete random variable, z^{X}: $\mathsf{E}\left(z^{\mathsf{X}}\right)$ is called the moment generating function (MGF), and also the probability generating function (PGF).

$$\mathcal{F}_{\mathsf{X}}(z) := \mathsf{E}\left(z^{\mathsf{X}}\right) = \sum_i z^i \, p_{\mathsf{X}}(i)$$

- If X is a continuous random variable, z^{X}: $\mathsf{E}\left(e^{j\omega\mathsf{X}}\right)$ is called the characteristic function. It is also the MGF.

$$\phi_{\mathsf{X}}(j\omega) := \mathsf{E}\left(e^{j\omega\mathsf{X}}\right) = \int_{-\infty}^{\infty} e^{j\omega x} f_{\mathsf{X}}(x)\, dx$$

  Often we use $j\omega = s$. Thus

$$\phi_{\mathsf{X}}(s) := \mathsf{E}\left(e^{s\mathsf{X}}\right) = \int_{-\infty}^{\infty} e^{sx} f_{\mathsf{X}}(x)\, dx$$

Moment and Probability Generating Functions

- For discrete random variables, recall the definition

$$\mathcal{F}_{\mathsf{X}}(z) := \mathsf{E}\left(z^{\mathsf{X}}\right) = \sum_i z^i p_{\mathsf{X}}(i).$$

- Evaluating $\mathcal{F}_{\mathsf{X}}(z)$ at $z = 0, 1$

$$\begin{aligned} \mathcal{F}_{\mathsf{X}}(0) &= \mathsf{Pr}(\mathsf{X} = 0) = p_0 \\ \mathcal{F}_{\mathsf{X}}(1) &= \sum_i p_{\mathsf{X}}(i) = 1 \end{aligned}$$

- Take the derivative of $\mathcal{F}_{\mathsf{X}}(z)$ w.r.t. z and evaluating at 0 and 1

$$\mathcal{F}'_{\mathsf{X}}(z) = \frac{d}{dz}\left(\sum_i z^i \, p_{\mathsf{X}}(i)\right) = \sum_i i z^{i-1} \, p_{\mathsf{X}}(i)$$

- $\mathcal{F}'_{\mathsf{X}}(0) = p_{\mathsf{X}}(1)$ $\qquad$ $\mathcal{F}'_{\mathsf{X}}(1) = \mathsf{E}(\mathsf{X})$

Moment and Probability Generating Functions

- For discrete random variables, recall the definition

$$\mathcal{F}_{\mathsf{X}}(z) := \mathsf{E}\left(z^{\mathsf{X}}\right) = \sum_i z^i p_{\mathsf{X}}(i).$$

- Evaluating $\mathcal{F}_{\mathsf{X}}(z)$ at $z = 0, 1$

$$\begin{aligned} \mathcal{F}_{\mathsf{X}}(0) &= \mathsf{Pr}(\mathsf{X} = 0) = p_0 \\ \mathcal{F}_{\mathsf{X}}(1) &= \sum_i p_{\mathsf{X}}(i) = 1 \end{aligned}$$

- Take the derivative of $\mathcal{F}_{\mathsf{X}}(z)$ w.r.t. z and evaluating at 0 and 1

$$\mathcal{F}'_{\mathsf{X}}(z) = \frac{d}{dz}\left(\sum_i z^i p_{\mathsf{X}}(i)\right) = \sum_i i z^{i-1} p_{\mathsf{X}}(i)$$

- $\mathcal{F}'_{\mathsf{X}}(0) = p_{\mathsf{X}}(1)$ $\qquad$ $\mathcal{F}'_{\mathsf{X}}(1) = \mathsf{E}(\mathsf{X})$

Moment and Probability Generating Functions

- More generally

$$\frac{\mathcal{F}_{\mathsf{X}}^{(i)}(0)}{i!} = p_{\mathsf{X}}(i)$$

$$\mathcal{F}_{\mathsf{X}}^{(k)}(1) = \mathsf{E}(\mathsf{X}(\mathsf{X}-1)(\mathsf{X}-2)\ldots(\mathsf{X}-k+1))$$

The last LHS is also called the k-th factorial moment.

Moments from MGF of Continuous RVs

- First moment from the Moment Generating Function

$$\phi'_X(s) = \frac{d}{ds}\left(\int_{-\infty}^{\infty} e^{sx} f_X(x)\, dx\right) = \int_{-\infty}^{\infty} x e^{sx} f_X(x)\, dx$$

$$\phi'_X(0) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \mathsf{E}(\mathsf{X})$$

- Second moment from the Moment Generating Function

$$\phi''_X(s) = \frac{d^2}{ds^2}\left(\int_{-\infty}^{\infty} e^{sx} f_X(x)\, dx\right) = \int_{-\infty}^{\infty} x^2 e^{sx} f_X(x)\, dx$$

$$\phi''_X(0) = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx = \mathsf{E}(\mathsf{X}^2)$$

- In general

$$\phi^{(n)}_X(0) = \int_{-\infty}^{\infty} x^n f_X(x)\, dx = \mathsf{E}(\mathsf{X}^n)$$

Moments from MGF of Continuous RVs

- First moment from the Moment Generating Function

$$\phi'_X(s) = \frac{d}{ds}\left(\int_{-\infty}^{\infty} e^{sx} f_X(x)\, dx\right) = \int_{-\infty}^{\infty} xe^{sx} f_X(x)\, dx$$

$$\phi'_X(0) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \mathsf{E}(X)$$

- Second moment from the Moment Generating Function

$$\phi''_X(s) = \frac{d^2}{ds^2}\left(\int_{-\infty}^{\infty} e^{sx} f_X(x)\, dx\right) = \int_{-\infty}^{\infty} x^2 e^{sx} f_X(x)\, dx$$

$$\phi''_X(0) = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx = \mathsf{E}(X^2)$$

- In general

$$\phi^{(n)}_X(0) = \int_{-\infty}^{\infty} x^n f_X(x)\, dx = \mathsf{E}(X^n)$$

Moments from MGF of Continuous RVs

- First moment from the Moment Generating Function

$$\begin{aligned}
\phi'_{\mathsf{X}}(s) &= \frac{d}{ds}\left(\int_{-\infty}^{\infty} e^{sx}\, f_{\mathsf{X}}(x)\, dx\right) = \int_{-\infty}^{\infty} xe^{sx}\, f_{\mathsf{X}}(x)\, dx \\
\phi'_{\mathsf{X}}(0) &= \int_{-\infty}^{\infty} x\, f_{\mathsf{X}}(x)\, dx = \mathsf{E}(\mathsf{X})
\end{aligned}$$

- Second moment from the Moment Generating Function

$$\begin{aligned}
\phi''_{\mathsf{X}}(s) &= \frac{d^2}{ds^2}\left(\int_{-\infty}^{\infty} e^{sx}\, f_{\mathsf{X}}(x)\, dx\right) = \int_{-\infty}^{\infty} x^2 e^{sx}\, f_{\mathsf{X}}(x)\, dx \\
\phi''_{\mathsf{X}}(0) &= \int_{-\infty}^{\infty} x^2\, f_{\mathsf{X}}(x)\, dx = \mathsf{E}(\mathsf{X}^2)
\end{aligned}$$

- In general

$$\phi^{(n)}_{\mathsf{X}}(0) = \int_{-\infty}^{\infty} x^n\, f_{\mathsf{X}}(x)\, dx = \mathsf{E}(\mathsf{X}^n)$$

Moments from MGF of Continuous RVs

- First moment from the Moment Generating Function

$$\phi'_X(s) = \frac{d}{ds}\left(\int_{-\infty}^{\infty} e^{sx} f_X(x)\, dx\right) = \int_{-\infty}^{\infty} x e^{sx} f_X(x)\, dx$$

$$\phi'_X(0) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \mathsf{E}(\mathsf{X})$$

- Second moment from the Moment Generating Function

$$\phi''_X(s) = \frac{d^2}{ds^2}\left(\int_{-\infty}^{\infty} e^{sx} f_X(x)\, dx\right) = \int_{-\infty}^{\infty} x^2 e^{sx} f_X(x)\, dx$$

$$\phi''_X(0) = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx = \mathsf{E}(\mathsf{X}^2)$$

- In general

$$\phi^{(n)}_X(0) = \int_{-\infty}^{\infty} x^n f_X(x)\, dx = \mathsf{E}(\mathsf{X}^n)$$

Moments from MGF of Continuous RVs

- First moment from the Moment Generating Function

$$\begin{aligned}
\phi'_{\mathsf{X}}(s) &= \frac{d}{ds}\left(\int_{-\infty}^{\infty} e^{sx} f_{\mathsf{X}}(x)\, dx\right) = \int_{-\infty}^{\infty} x e^{sx} f_{\mathsf{X}}(x)\, dx \\
\phi'_{\mathsf{X}}(0) &= \int_{-\infty}^{\infty} x\, f_{\mathsf{X}}(x)\, dx = \mathsf{E}(\mathsf{X})
\end{aligned}$$

- Second moment from the Moment Generating Function

$$\begin{aligned}
\phi''_{\mathsf{X}}(s) &= \frac{d^2}{ds^2}\left(\int_{-\infty}^{\infty} e^{sx} f_{\mathsf{X}}(x)\, dx\right) = \int_{-\infty}^{\infty} x^2 e^{sx} f_{\mathsf{X}}(x)\, dx \\
\phi''_{\mathsf{X}}(0) &= \int_{-\infty}^{\infty} x^2\, f_{\mathsf{X}}(x)\, dx = \mathsf{E}(\mathsf{X}^2)
\end{aligned}$$

- In general

$$\phi^{(n)}_{\mathsf{X}}(0) = \int_{-\infty}^{\infty} x^n\, f_{\mathsf{X}}(x)\, dx = \mathsf{E}(\mathsf{X}^n)$$

Two or More Random Variables

- On the same sample space we can define multiple random variables; begin with two, say X and Y.
- Like joint probability of two events, can define joint pmf.
- Example: Roll a fair die till you get a 1. Let the total number rolls be the random variable X. Let the number of 2s before the first 1 be the random variable Y. Let the number of evens before the first 1, be the random variable Z. Verify the following conditional and joint pmfs.

$$
\begin{aligned}
p_{\mathsf{Z}|\mathsf{X}}(k|i) &= \binom{i-1}{k}(3/5)^k(2/5)^{i-k} \qquad \text{for } k = 0, 1, \ldots i-1 \\
p_{\mathsf{XY}}(i,j) &= \binom{i-1}{j}(1/5)^j(4/5)^{i-1-j}\,(5/6)^{i-1}(1/6) \\
&\qquad \text{for } i = 1, 2, \ldots \text{ and } j = 0, 1, \ldots i-1 \\
p_{\mathsf{XZ}}(i,k) &= \binom{i-1}{k}(3/5)^k(2/5)^{i-k}\,(5/6)^{i-1}(1/6) \\
&\qquad \text{for } i = 1, 2, \ldots \text{ and } k = 0, 1, \ldots i-1
\end{aligned}
$$

Two or More Random Variables

- On the same sample space we can define multiple random variables; begin with two, say X and Y.
- Like joint probability of two events, can define joint pmf.
- Example: Roll a fair die till you get a 1. Let the total number rolls be the random variable X. Let the number of 2s before the first 1 be the random variable Y. Let the number of evens before the first 1, be the random variable Z. Verify the following conditional and joint pmfs.

$$
\begin{aligned}
p_{\mathsf{Z}|\mathsf{X}}(k|i) &= \binom{i-1}{k}(3/5)^k(2/5)^{i-k} \qquad \text{for } k = 0, 1, \ldots i-1 \\
p_{\mathsf{XY}}(i,j) &= \binom{i-1}{j}(1/5)^j(4/5)^{i-1-j}\,(5/6)^{i-1}(1/6) \\
&\qquad \text{for } i = 1, 2, \ldots \text{ and } j = 0, 1, \ldots i-1 \\
p_{\mathsf{XZ}}(i,k) &= \binom{i-1}{k}(3/5)^k(2/5)^{i-k}\,(5/6)^{i-1}(1/6) \\
&\qquad \text{for } i = 1, 2, \ldots \text{ and } k = 0, 1, \ldots i-1
\end{aligned}
$$

Two or More Random Variables

- On the same sample space we can define multiple random variables; begin with two, say X and Y.
- Like joint probability of two events, can define joint pmf.
- Example: Roll a fair die till you get a 1. Let the total number rolls be the random variable X. Let the number of 2s before the first 1 be the random variable Y. Let the number of evens before the first 1, be the random variable Z. Verify the following conditional and joint pmfs.

$$
\begin{aligned}
p_{\mathsf{Z}|\mathsf{X}}(k|i) &= \binom{i-1}{k}(3/5)^k(2/5)^{i-k} \qquad \text{for } k = 0, 1, \ldots i-1 \\
p_{\mathsf{XY}}(i,j) &= \binom{i-1}{j}(1/5)^j(4/5)^{i-1-j}\,(5/6)^{i-1}(1/6) \\
&\qquad \text{for } i = 1, 2, \ldots \text{ and } j = 0, 1, \ldots i-1 \\
p_{\mathsf{XZ}}(i,k) &= \binom{i-1}{k}(3/5)^k(2/5)^{i-k}\,(5/6)^{i-1}(1/6) \\
&\qquad \text{for } i = 1, 2, \ldots \text{ and } k = 0, 1, \ldots i-1
\end{aligned}
$$

Two or More Random Variables

- On the same sample space we can define multiple random variables; begin with two, say X and Y.
- Like joint probability of two events, can define joint pmf.
- Example: Roll a fair die till you get a 1. Let the total number rolls be the random variable X. Let the number of 2s before the first 1 be the random variable Y. Let the number of evens before the first 1, be the random variable Z. Verify the following conditional and joint pmfs.

$$
\begin{aligned}
p_{\mathsf{Z}|\mathsf{X}}(k|i) &= \binom{i-1}{k}(3/5)^k(2/5)^{i-k} \qquad \text{for } k = 0, 1, \ldots i-1 \\
p_{\mathsf{XY}}(i,j) &= \binom{i-1}{j}(1/5)^j(4/5)^{i-1-j}\,(5/6)^{i-1}(1/6) \\
& \qquad \text{for } i = 1, 2, \ldots \text{ and } j = 0, 1, \ldots i-1 \\
p_{\mathsf{XZ}}(i,k) &= \binom{i-1}{k}(3/5)^k(2/5)^{i-k}\,(5/6)^{i-1}(1/6) \\
& \qquad \text{for } i = 1, 2, \ldots \text{ and } k = 0, 1, \ldots i-1
\end{aligned}
$$

Two or More Random Variables

- More generally, like the joint probability of two events and the joint pmf, can define joint cdf for X and Y.

$$
\begin{aligned}
F_{\mathsf{X},\mathsf{Y}}(x, y) &:= \Pr(\mathsf{X} \le x, \mathsf{Y} \le y) \\
f_{\mathsf{X},\mathsf{Y}}(x, y) &:= \frac{\partial^2 F_{\mathsf{X},\mathsf{Y}}(x, y)}{\partial x \, \partial y}
\end{aligned}
$$

- Informally, $f_{\mathsf{X},\mathsf{Y}}(x, y)\, dx\, dy$ is the probability that X and Y lie in the infinitesimal area of $dx\, dy$ near (x, y)
- Some easy identities.

$$
\begin{aligned}
F_{\mathsf{X}}(x) &= F_{\mathsf{XY}}(x, \infty) \\
F_{\mathsf{Y}}(y) &= F_{\mathsf{XY}}(\infty, y) \\
\Pr(x_1 \le \mathsf{X} \le x_2, y_1 \le \mathsf{Y} \le y_2) &= F_{\mathsf{XY}}(x_2, y_2) - F_{\mathsf{XY}}(x_1, y_2) - \\
&\quad F_{\mathsf{XY}}(x_2, y_1) + F_{\mathsf{XY}}(x_1, y_1)
\end{aligned}
$$

Two or More Random Variables

- More generally, like the joint probability of two events and the joint pmf, can define joint cdf for X and Y.

$$
\begin{aligned}
F_{\mathsf{X},\mathsf{Y}}(x,y) &:= \Pr(\mathsf{X} \le x, \mathsf{Y} \le y) \\
f_{\mathsf{X},\mathsf{Y}}(x,y) &:= \frac{\partial^2 F_{\mathsf{X},\mathsf{Y}}(x,y)}{\partial x \, \partial y}
\end{aligned}
$$

- Informally, $f_{\mathsf{X},\mathsf{Y}}(x,y)\, dx\, dy$ is the probability that X and Y lie in the infinitesimal area of $dx\, dy$ near (x,y)
- Some easy identities.

$$
\begin{aligned}
F_{\mathsf{X}}(x) &= F_{\mathsf{XY}}(x,\infty) \\
F_{\mathsf{Y}}(y) &= F_{\mathsf{XY}}(\infty,y) \\
\Pr(x_1 \le \mathsf{X} \le x_2, y_1 \le \mathsf{Y} \le y_2) &= F_{\mathsf{XY}}(x_2,y_2) - F_{\mathsf{XY}}(x_1,y_2) - \\
&\quad F_{\mathsf{XY}}(x_2,y_1) + F_{\mathsf{XY}}(x_1,y_1)
\end{aligned}
$$

Two or More Random Variables

- More generally, like the joint probability of two events and the joint pmf, can define joint cdf for X and Y.

$$\begin{aligned} F_{\mathsf{X},\mathsf{Y}}(x,y) &:= \mathsf{Pr}(\mathsf{X} \leq x, \mathsf{Y} \leq y) \\ f_{\mathsf{X},\mathsf{Y}}(x,y) &:= \frac{\partial^2 F_{\mathsf{X},\mathsf{Y}}(x,y)}{\partial x\, \partial y} \end{aligned}$$

- Informally, $f_{\mathsf{X},\mathsf{Y}}(x,y)\, dx\, dy$ is the probability that X and Y lie in the infinitesimal area of $dx\, dy$ near (x,y)
- Some easy identities.

$$\begin{aligned} F_{\mathsf{X}}(x) &= F_{\mathsf{XY}}(x,\infty) \\ F_{\mathsf{Y}}(y) &= F_{\mathsf{XY}}(\infty,y) \\ \mathsf{Pr}(x_1 \leq \mathsf{X} \leq x_2, y_1 \leq \mathsf{Y} \leq y_2) &= F_{\mathsf{XY}}(x_2,y_2) - F_{\mathsf{XY}}(x_1,y_2) - \\ &\quad F_{\mathsf{XY}}(x_2,y_1) + F_{\mathsf{XY}}(x_1,y_1) \end{aligned}$$

Two or More Random Variables

- More identities

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$$

$$\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = f_X(x)$$

- X and Y are independent if

$$F_{XY}(x,y) = F_X(x)F_Y(y)$$

This also implies

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

- More identities

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$$
$$\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = f_X(x)$$

- X and Y are independent if

$$F_{XY}(x,y) = F_X(x)F_Y(y)$$

This also implies

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

Two or More Random Variables

- More identities

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathsf{XY}}(x,y)\,dx\,dy = 1$$
$$\int_{-\infty}^{\infty} f_{\mathsf{XY}}(x,y)\,dy = f_{\mathsf{X}}(x)$$

- X and Y are independent if

$$F_{\mathsf{XY}}(x,y) = F_{\mathsf{X}}(x)F_{\mathsf{Y}}(y)$$

This also implies

$$f_{\mathsf{XY}}(x,y) = f_{\mathsf{X}}(x)f_{\mathsf{Y}}(y)$$

Two or More Random Variables

- More identities

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathsf{XY}}(x,y)\,dx\,dy = 1$$
$$\int_{-\infty}^{\infty} f_{\mathsf{XY}}(x,y)\,dy = f_{\mathsf{X}}(x)$$

- X and Y are independent if

$$F_{\mathsf{XY}}(x,y) = F_{\mathsf{X}}(x)F_{\mathsf{Y}}(y)$$

This also implies

$$f_{\mathsf{XY}}(x,y) = f_{\mathsf{X}}(x)f_{\mathsf{Y}}(y)$$

Two or More Random Variables: Tutorial Problems

1. Obtain the distribution of $\mathsf{Z} = \min(\mathsf{X}, \mathsf{Y})$
   Key observation: $\{\mathsf{Z} > z\} \implies \{\mathsf{X} > z, \mathsf{Y} > z\})$
2. Obtain the distribution of $\mathsf{Z} = \max(\mathsf{X}, \mathsf{Y})$
   Key observation: $\{\mathsf{Z} \leq z\} \implies \{\mathsf{X} \leq z, \mathsf{Y} \leq z\})$
3. Order statistics of N iid variables: Take the N realisations of the random variables and sort them in, say increasing order. The r-th element of the sorted list is called the r-th order statistics.
   Use the results from either of the preceding examples.
4. Obtain the distribution of $\mathsf{Z} = \mathsf{X} + \mathsf{Y}$.

Expectations of Functions of Two Random Variables

- Let $\mathsf{Z} = g(\mathsf{X}, \mathsf{Y})$. Using arguments identical to the mean of a function of a random variable, we can show

$$\mathsf{E}(\mathsf{X}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{\mathsf{XY}}(x, y) dxdy$$

-

$$\begin{aligned}
\mathsf{E}(\mathsf{X}+\mathsf{Y}) &= \int \int (x+y) f_{\mathsf{X},\mathsf{Y}}(x,y)\, dx\, dy \\
&= \int \int x f_{\mathsf{X},\mathsf{Y}}(x,y)\, dx\, dy + \int \int y f_{\mathsf{X},\mathsf{Y}}(x,y)\, dx\, dy \\
&= \int x \int f_{\mathsf{X},\mathsf{Y}}(x,y)\, dy\, dx + \int y \int f_{\mathsf{X},\mathsf{Y}}(x,y)\, dx\, dy \\
&= \int x f_{\mathsf{X}}(x)\, dx + \int y f_{\mathsf{Y}}(y)\, dy \\
&= \mathsf{E}(\mathsf{X}) + \mathsf{E}(\mathsf{Y})
\end{aligned}$$

Observe that there were NO assumptions on X and Y.

Expectations of Functions of Two Random Variables

- Let $\mathsf{Z} = g(\mathsf{X}, \mathsf{Y})$. Using arguments identical to the mean of a function of a random variable, we can show

$$\mathsf{E}(\mathsf{X}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{\mathsf{XY}}(x, y) dx dy$$

-

$$\begin{aligned}
\mathsf{E}(\mathsf{X}+\mathsf{Y}) &= \int\int (x+y)\, f_{\mathsf{X},\mathsf{Y}}(x,y)\, dx\, dy \\
&= \int\int x\, f_{\mathsf{X},\mathsf{Y}}(x,y)\, dx\, dy + \int\int y\, f_{\mathsf{X},\mathsf{Y}}(x,y)\, dx\, dy \\
&= \int x \int f_{\mathsf{X},\mathsf{Y}}(x,y)\, dy\, dx + \int y \int f_{\mathsf{X},\mathsf{Y}}(x,y)\, dx\, dy \\
&= \int x\, f_{\mathsf{X}}(x)\, dx + \int y\, f_{\mathsf{Y}}(y)\, dy \\
&= \mathsf{E}(\mathsf{X}) + \mathsf{E}(\mathsf{Y})
\end{aligned}$$

Observe that there were NO assumptions on X and Y.

Expectations of Functions of Two Random Variables

- Let $\mathsf{Z} = g(\mathsf{X}, \mathsf{Y})$. Using arguments identical to the mean of a function of a random variable, we can show

$$\mathsf{E}(\mathsf{X}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{\mathsf{XY}}(x, y) dx dy$$

-

$$\begin{aligned}
\mathsf{E}(\mathsf{X} + \mathsf{Y}) &= \int \int (x + y)\, f_{\mathsf{X},\mathsf{Y}}(x, y)\ dx\, dy \\
&= \int \int x\, f_{\mathsf{X},\mathsf{Y}}(x, y)\ dx\, dy + \int \int y\, f_{\mathsf{X},\mathsf{Y}}(x, y)\ dx\, dy \\
&= \int x \int f_{\mathsf{X},\mathsf{Y}}(x, y)\ dy\, dx + \int y \int f_{\mathsf{X},\mathsf{Y}}(x, y)\ dx\, dy \\
&= \int x\, f_{\mathsf{X}}(x)\ dx + \int y\, f_{\mathsf{Y}}(y)\ dy \\
&= \mathsf{E}(\mathsf{X}) + \mathsf{E}(\mathsf{Y})
\end{aligned}$$

Observe that there were NO assumptions on X and Y.

Expectations of Functions of Two Random Variables

- Let $\mathsf{Z} = g(\mathsf{X}, \mathsf{Y})$. Using arguments identical to the mean of a function of a random variable, we can show

$$\mathsf{E}(\mathsf{X}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{\mathsf{XY}}(x,y)\,dx\,dy$$

-

$$\begin{aligned}
\mathsf{E}(\mathsf{X}+\mathsf{Y}) &= \int\int (x+y)\, f_{\mathsf{X},\mathsf{Y}}(x,y)\; dx\, dy \\
&= \int\int x\, f_{\mathsf{X},\mathsf{Y}}(x,y)\; dx\, dy + \int\int y\, f_{\mathsf{X},\mathsf{Y}}(x,y)\; dx\, dy \\
&= \int x \int f_{\mathsf{X},\mathsf{Y}}(x,y)\; dy\, dx + \int y \int f_{\mathsf{X},\mathsf{Y}}(x,y)\; dx\, dy \\
&= \int x\, f_{\mathsf{X}}(x)\; dx + \int y\, f_{\mathsf{Y}}(y)\; dy \\
&= \mathsf{E}(\mathsf{X}) + \mathsf{E}(\mathsf{Y})
\end{aligned}$$

Observe that there were NO assumptions on X and Y.

Expectations of Functions of Two Random Variables

- Let $\mathsf{Z} = g(\mathsf{X}, \mathsf{Y})$. Using arguments identical to the mean of a function of a random variable, we can show

$$\mathsf{E}(\mathsf{X}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{\mathsf{XY}}(x, y) dx dy$$

-

$$\begin{aligned}
\mathsf{E}(\mathsf{X} + \mathsf{Y}) &= \int \int (x + y)\, f_{\mathsf{X},\mathsf{Y}}(x, y)\, dx\, dy \\
&= \int \int x\, f_{\mathsf{X},\mathsf{Y}}(x, y)\, dx\, dy + \int \int y\, f_{\mathsf{X},\mathsf{Y}}(x, y)\, dx\, dy \\
&= \int x \int f_{\mathsf{X},\mathsf{Y}}(x, y)\, dy\, dx + \int y \int f_{\mathsf{X},\mathsf{Y}}(x, y)\, dx\, dy \\
&= \int x\, f_{\mathsf{X}}(x)\, dx + \int y\, f_{\mathsf{Y}}(y)\, dy \\
&= \mathsf{E}(\mathsf{X}) + \mathsf{E}(\mathsf{Y})
\end{aligned}$$

Observe that there were NO assumptions on X and Y.

Expectations of Functions of Two Random Variables

- Let $\mathsf{Z} = g(\mathsf{X}, \mathsf{Y})$. Using arguments identical to the mean of a function of a random variable, we can show

$$\mathsf{E}(\mathsf{X}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{\mathsf{XY}}(x,y)\,dx dy$$

-

$$\begin{aligned}
\mathsf{E}(\mathsf{X}+\mathsf{Y}) &= \int\int (x+y)\, f_{\mathsf{X},\mathsf{Y}}(x,y)\; dx\, dy \\
&= \int\int x\, f_{\mathsf{X},\mathsf{Y}}(x,y)\; dx\, dy + \int\int y\, f_{\mathsf{X},\mathsf{Y}}(x,y)\; dx\, dy \\
&= \int x \int f_{\mathsf{X},\mathsf{Y}}(x,y)\; dy\, dx + \int y \int f_{\mathsf{X},\mathsf{Y}}(x,y)\; dx\, dy \\
&= \int x\, f_{\mathsf{X}}(x)\; dx + \int y\, f_{\mathsf{Y}}(y)\; dy \\
&= \mathsf{E}(\mathsf{X}) + \mathsf{E}(\mathsf{Y})
\end{aligned}$$

Observe that there were NO assumptions on X and Y.

Covariance of two Random Variables

- Covariance is a measure of two random variables, say X and Y are related. How they vary 'together.'
- **Informally**
  - Want to know if high values of X means high values of Y **and** low values of X means low values of Y.
  - Could be the opposite: High values of X means low values of Y **and** low values of X means high values of Y.
  - They could 'balance out': high values of X could give rise to both high and low values of Y. Similarly for low values.
  - We now define the notions of positive correlation, negative correlation and uncorrelatedness between two random variables.

Covariance of two Random Variables

- For convenience we will write $\mathsf{E}(\mathsf{X}) = \mu_{\mathsf{X}}$ and $\mathsf{E}(\mathsf{Y}) = \mu_{\mathsf{Y}}$.
- Denote covariance of X and Y by $\mathsf{COV}(\mathsf{X}, \mathsf{Y})$ and define

$$\mathsf{COV}(\mathsf{X}, \mathsf{Y}) = \mathsf{E}((\mathsf{X} - \mu_{\mathsf{X}})\ (\mathsf{Y} - \mu_{\mathsf{Y}}))$$

Thus

$$\begin{aligned}
\mathsf{COV}(\mathsf{X}, \mathsf{Y}) &= \mathsf{E}((\mathsf{X} - \mu_{\mathsf{X}})\ (\mathsf{Y} - \mu_{\mathsf{Y}})) \\
&= \mathsf{E}(\mathsf{XY} - \mu_{\mathsf{X}}\mathsf{Y} - \mathsf{X}\mu_{\mathsf{Y}} + \mu_{\mathsf{X}}\mu_{\mathsf{Y}}) \\
&= \mathsf{E}(\mathsf{XY}) - \mathsf{E}(\mu_{\mathsf{X}}\mathsf{Y}) - \mathsf{E}(\mathsf{X}\mu_{\mathsf{Y}}) + \mathsf{E}(\mu_{\mathsf{X}}\mu_{\mathsf{Y}}) \\
&= \mathsf{E}(\mathsf{XY}) - \mu_{\mathsf{X}}\mathsf{E}(\mathsf{Y}) - \mathsf{E}(\mathsf{X})\,\mu_{\mathsf{Y}} + \mu_{\mathsf{X}}\mu_{\mathsf{Y}} \\
&= \mathsf{E}(\mathsf{XY}) - \mu_{\mathsf{X}}\mu_{\mathsf{Y}} - \mu_{\mathsf{X}}\mu_{\mathsf{Y}} + \mu_{\mathsf{X}}\mu_{\mathsf{Y}} \\
&= \mathsf{E}(\mathsf{XY}) - \mu_{\mathsf{X}}\mu_{\mathsf{Y}}
\end{aligned}$$

- What is $\mathsf{COV}(\mathsf{X}, \mathsf{X})$?

Covariance of two Random Variables

- For convenience we will write $\mathsf{E}(\mathsf{X}) = \mu_{\mathsf{X}}$ and $\mathsf{E}(\mathsf{Y}) = \mu_{\mathsf{Y}}$.
- Denote covariance of X and Y by $\mathsf{COV}(\mathsf{X}, \mathsf{Y})$ and define

$$\mathsf{COV}(\mathsf{X}, \mathsf{Y}) = \mathsf{E}((\mathsf{X} - \mu_{\mathsf{X}})\ (\mathsf{Y} - \mu_{\mathsf{Y}}))$$

Thus

$$\begin{aligned} \mathsf{COV}(\mathsf{X}, \mathsf{Y}) &= \mathsf{E}((\mathsf{X} - \mu_{\mathsf{X}})\ (\mathsf{Y} - \mu_{\mathsf{Y}}))\text{‘} \\ &= \mathsf{E}(\mathsf{XY} - \mu_{\mathsf{X}}\mathsf{Y} - \mathsf{X}\mu_{\mathsf{Y}} + \mu_{\mathsf{X}}\mu_{\mathsf{Y}}) \\ &= \mathsf{E}(\mathsf{XY}) - \mathsf{E}(\mu_{\mathsf{X}}\mathsf{Y}) - \mathsf{E}(\mathsf{X}\mu_{\mathsf{Y}}) + \mathsf{E}(\mu_{\mathsf{X}}\mu_{\mathsf{Y}}) \\ &= \mathsf{E}(\mathsf{XY}) - \mu_{\mathsf{X}}\mathsf{E}(\mathsf{Y}) - \mathsf{E}(\mathsf{X})\,\mu_{\mathsf{Y}} + \mu_{\mathsf{X}}\mu_{\mathsf{Y}} \\ &= \mathsf{E}(\mathsf{XY}) - \mu_{\mathsf{X}}\mu_{\mathsf{Y}} - \mu_{\mathsf{X}}\mu_{\mathsf{Y}} + \mu_{\mathsf{X}}\mu_{\mathsf{Y}} \\ &= \mathsf{E}(\mathsf{XY}) - \mu_{\mathsf{X}}\mu_{\mathsf{Y}} \end{aligned}$$

- What is $\mathsf{COV}(\mathsf{X}, \mathsf{X})$?

Covariance of two Random Variables

- For convenience we will write $\mathsf{E}(\mathsf{X}) = \mu_\mathsf{X}$ and $\mathsf{E}(\mathsf{Y}) = \mu_\mathsf{Y}$.
- Denote covariance of X and Y by $\mathsf{COV}(\mathsf{X},\mathsf{Y})$ and define

$$\mathsf{COV}(\mathsf{X},\mathsf{Y}) = \mathsf{E}((\mathsf{X}-\mu_\mathsf{X})\ (\mathsf{Y}-\mu_\mathsf{Y}))$$

Thus

$$\begin{aligned}
\mathsf{COV}(\mathsf{X},\mathsf{Y}) &= \mathsf{E}((\mathsf{X}-\mu_\mathsf{X})\ (\mathsf{Y}-\mu_\mathsf{Y}))\text{`} \\
&= \mathsf{E}(\mathsf{XY} - \mu_\mathsf{X}\mathsf{Y} - \mathsf{X}\mu_\mathsf{Y} + \mu_\mathsf{X}\mu_\mathsf{Y}) \\
&= \mathsf{E}(\mathsf{XY}) - \mathsf{E}(\mu_\mathsf{X}\mathsf{Y}) - \mathsf{E}(\mathsf{X}\mu_\mathsf{Y}) + \mathsf{E}(\mu_\mathsf{X}\mu_\mathsf{Y}) \\
&= \mathsf{E}(\mathsf{XY}) - \mu_\mathsf{X}\mathsf{E}(\mathsf{Y}) - \mathsf{E}(\mathsf{X})\,\mu_\mathsf{Y} + \mu_\mathsf{X}\mu_\mathsf{Y} \\
&= \mathsf{E}(\mathsf{XY}) - \mu_\mathsf{X}\mu_\mathsf{Y} - \mu_\mathsf{X}\mu_\mathsf{Y} + \mu_\mathsf{X}\mu_\mathsf{Y} \\
&= \mathsf{E}(\mathsf{XY}) - \mu_\mathsf{X}\mu_\mathsf{Y}
\end{aligned}$$

- What is $\mathsf{COV}(\mathsf{X},\mathsf{X})$?

Covariance of two Random Variables

- For convenience we will write $\mathsf{E}(\mathsf{X}) = \mu_{\mathsf{X}}$ and $\mathsf{E}(\mathsf{Y}) = \mu_{\mathsf{Y}}$.
- Denote covariance of X and Y by $\mathsf{COV}(\mathsf{X}, \mathsf{Y})$ and define

$$\mathsf{COV}(\mathsf{X}, \mathsf{Y}) = \mathsf{E}((\mathsf{X} - \mu_{\mathsf{X}})\ (\mathsf{Y} - \mu_{\mathsf{Y}}))$$

Thus

$$\begin{aligned}
\mathsf{COV}(\mathsf{X}, \mathsf{Y}) &= \mathsf{E}((\mathsf{X} - \mu_{\mathsf{X}})\ (\mathsf{Y} - \mu_{\mathsf{Y}}))` \\
&= \mathsf{E}(\mathsf{XY} - \mu_{\mathsf{X}}\mathsf{Y} - \mathsf{X}\mu_{\mathsf{Y}} + \mu_{\mathsf{X}}\mu_{\mathsf{Y}}) \\
&= \mathsf{E}(\mathsf{XY}) - \mathsf{E}(\mu_{\mathsf{X}}\mathsf{Y}) - \mathsf{E}(\mathsf{X}\mu_{\mathsf{Y}}) + \mathsf{E}(\mu_{\mathsf{X}}\mu_{\mathsf{Y}}) \\
&= \mathsf{E}(\mathsf{XY}) - \mu_{\mathsf{X}}\mathsf{E}(\mathsf{Y}) - \mathsf{E}(\mathsf{X})\,\mu_{\mathsf{Y}} + \mu_{\mathsf{X}}\mu_{\mathsf{Y}} \\
&= \mathsf{E}(\mathsf{XY}) - \mu_{\mathsf{X}}\mu_{\mathsf{Y}} - \mu_{\mathsf{X}}\mu_{\mathsf{Y}} + \mu_{\mathsf{X}}\mu_{\mathsf{Y}} \\
&= \mathsf{E}(\mathsf{XY}) - \mu_{\mathsf{X}}\mu_{\mathsf{Y}}
\end{aligned}$$

- What is $\mathsf{COV}(\mathsf{X}, \mathsf{X})$?

Covariance of two Random Variables

- For convenience we will write $\mathsf{E}(\mathsf{X}) = \mu_\mathsf{X}$ and $\mathsf{E}(\mathsf{Y}) = \mu_\mathsf{Y}$.
- Denote covariance of X and Y by $\mathsf{COV}(\mathsf{X},\mathsf{Y})$ and define

$$\mathsf{COV}(\mathsf{X},\mathsf{Y}) = \mathsf{E}((\mathsf{X}-\mu_\mathsf{X})\ (\mathsf{Y}-\mu_\mathsf{Y}))$$

Thus

$$\begin{aligned}
\mathsf{COV}(\mathsf{X},\mathsf{Y}) &= \mathsf{E}((\mathsf{X}-\mu_\mathsf{X})\ (\mathsf{Y}-\mu_\mathsf{Y}))` \\
&= \mathsf{E}(\mathsf{XY} - \mu_\mathsf{X}\mathsf{Y} - \mathsf{X}\mu_\mathsf{Y} + \mu_\mathsf{X}\mu_\mathsf{Y}) \\
&= \mathsf{E}(\mathsf{XY}) - \mathsf{E}(\mu_\mathsf{X}\mathsf{Y}) - \mathsf{E}(\mathsf{X}\mu_\mathsf{Y}) + \mathsf{E}(\mu_\mathsf{X}\mu_\mathsf{Y}) \\
&= \mathsf{E}(\mathsf{XY}) - \mu_\mathsf{X}\mathsf{E}(\mathsf{Y}) - \mathsf{E}(\mathsf{X})\,\mu_\mathsf{Y} + \mu_\mathsf{X}\mu_\mathsf{Y} \\
&= \mathsf{E}(\mathsf{XY}) - \mu_\mathsf{X}\mu_\mathsf{Y} - \mu_\mathsf{X}\mu_\mathsf{Y} + \mu_\mathsf{X}\mu_\mathsf{Y} \\
&= \mathsf{E}(\mathsf{XY}) - \mu_\mathsf{X}\mu_\mathsf{Y}
\end{aligned}$$

- What is $\mathsf{COV}(\mathsf{X},\mathsf{X})$?

Covariance of two Random Variables

- For convenience we will write $\mathsf{E}(\mathsf{X}) = \mu_\mathsf{X}$ and $\mathsf{E}(\mathsf{Y}) = \mu_\mathsf{Y}$.
- Denote covariance of X and Y by $\mathsf{COV}(\mathsf{X}, \mathsf{Y})$ and define

$$\mathsf{COV}(\mathsf{X}, \mathsf{Y}) = \mathsf{E}((\mathsf{X} - \mu_\mathsf{X})\ (\mathsf{Y} - \mu_\mathsf{Y}))$$

Thus

$$\begin{aligned}
\mathsf{COV}(\mathsf{X}, \mathsf{Y}) &= \mathsf{E}((\mathsf{X} - \mu_\mathsf{X})\ (\mathsf{Y} - \mu_\mathsf{Y}))\text{'} \\
&= \mathsf{E}(\mathsf{XY} - \mu_\mathsf{X}\mathsf{Y} - \mathsf{X}\mu_\mathsf{Y} + \mu_\mathsf{X}\mu_\mathsf{Y}) \\
&= \mathsf{E}(\mathsf{XY}) - \mathsf{E}(\mu_\mathsf{X}\mathsf{Y}) - \mathsf{E}(\mathsf{X}\mu_\mathsf{Y}) + \mathsf{E}(\mu_\mathsf{X}\mu_\mathsf{Y}) \\
&= \mathsf{E}(\mathsf{XY}) - \mu_\mathsf{X}\mathsf{E}(\mathsf{Y}) - \mathsf{E}(\mathsf{X})\,\mu_\mathsf{Y} + \mu_\mathsf{X}\mu_\mathsf{Y} \\
&= \mathsf{E}(\mathsf{XY}) - \mu_\mathsf{X}\mu_\mathsf{Y} - \mu_\mathsf{X}\mu_\mathsf{Y} + \mu_\mathsf{X}\mu_\mathsf{Y} \\
&= \mathsf{E}(\mathsf{XY}) - \mu_\mathsf{X}\mu_\mathsf{Y}
\end{aligned}$$

- What is $\mathsf{COV}(\mathsf{X}, \mathsf{X})$?

Covariance of two Random Variables

- For convenience we will write $\mathsf{E}(\mathsf{X}) = \mu_{\mathsf{X}}$ and $\mathsf{E}(\mathsf{Y}) = \mu_{\mathsf{Y}}$.
- Denote covariance of X and Y by $\mathsf{COV}(\mathsf{X},\mathsf{Y})$ and define

$$\mathsf{COV}(\mathsf{X},\mathsf{Y}) = \mathsf{E}((\mathsf{X}-\mu_{\mathsf{X}})\ (\mathsf{Y}-\mu_{\mathsf{Y}}))$$

Thus

$$\begin{aligned}
\mathsf{COV}(\mathsf{X},\mathsf{Y}) &= \mathsf{E}((\mathsf{X}-\mu_{\mathsf{X}})\ (\mathsf{Y}-\mu_{\mathsf{Y}}))\text{`} \\
&= \mathsf{E}(\mathsf{XY} - \mu_{\mathsf{X}}\mathsf{Y} - \mathsf{X}\mu_{\mathsf{Y}} + \mu_{\mathsf{X}}\mu_{\mathsf{Y}}) \\
&= \mathsf{E}(\mathsf{XY}) - \mathsf{E}(\mu_{\mathsf{X}}\mathsf{Y}) - \mathsf{E}(\mathsf{X}\mu_{\mathsf{Y}}) + \mathsf{E}(\mu_{\mathsf{X}}\mu_{\mathsf{Y}}) \\
&= \mathsf{E}(\mathsf{XY}) - \mu_{\mathsf{X}}\mathsf{E}(\mathsf{Y}) - \mathsf{E}(\mathsf{X})\,\mu_{\mathsf{Y}} + \mu_{\mathsf{X}}\mu_{\mathsf{Y}} \\
&= \mathsf{E}(\mathsf{XY}) - \mu_{\mathsf{X}}\mu_{\mathsf{Y}} - \mu_{\mathsf{X}}\mu_{\mathsf{Y}} + \mu_{\mathsf{X}}\mu_{\mathsf{Y}} \\
&= \mathsf{E}(\mathsf{XY}) - \mu_{\mathsf{X}}\mu_{\mathsf{Y}}
\end{aligned}$$

- What is $\mathsf{COV}(\mathsf{X},\mathsf{X})$?

Covariance of two Independent Random Variables

Let X and Y be two independent random variables Want to know $\mathsf{E}(\mathsf{XY})$.

$$\begin{aligned}
\mathsf{E}(\mathsf{XY}) &= \sum_j \sum_k j\, k\, p_{\mathsf{X},\mathsf{Y}}(j,k) \\
&= \sum_j \sum_k j\, k\, p_{\mathsf{X}}(j)\, p_{\mathsf{Y}}(k) \\
&= \sum_j j\, p_{\mathsf{X}}(j) \sum_k k\, p_{\mathsf{Y}}(k) \\
&= \left(\sum_j j\, p_{\mathsf{X}}(j)\right)\left(\sum_k k\, p_{\mathsf{Y}}(k)\right) \\
&= \mathsf{E}(\mathsf{X})\ \mathsf{E}(\mathsf{Y})
\end{aligned}$$

$$\mathsf{COV}(\mathsf{X},\mathsf{Y}) = ??$$

Thus independence implies uncorrelatedness. Does uncorrelatedness imply independence?

Covariance of two Independent Random Variables

Let X and Y be two independent random variables Want to know $\mathsf{E}(\mathsf{XY})$.

$$
\begin{aligned}
\mathsf{E}(\mathsf{XY}) &= \sum_j \sum_k j\, k\, p_{\mathsf{X},\mathsf{Y}}(j,k) \\
&= \sum_j \sum_k j\, k\, p_{\mathsf{X}}(j)\, p_{\mathsf{Y}}(k) \\
&= \sum_j j\, p_{\mathsf{X}}(j) \sum_k k\, p_{\mathsf{Y}}(k) \\
&= \left(\sum_j j\, p_{\mathsf{X}}(j)\right)\left(\sum_k k\, p_{\mathsf{Y}}(k)\right) \\
&= \mathsf{E}(\mathsf{X})\ \mathsf{E}(\mathsf{Y})
\end{aligned}
$$

$$
\mathsf{COV}(\mathsf{X},\mathsf{Y}) = ??
$$

Thus independence implies uncorrelatedness. Does uncorrelatedness imply independence?

Covariance of two Independent Random Variables

Let X and Y be two independent random variables Want to know $\mathsf{E}(\mathsf{XY})$.

$$
\begin{aligned}
\mathsf{E}(\mathsf{XY}) &= \sum_j \sum_k j\, k\, p_{\mathsf{X},\mathsf{Y}}(j,k) \\
&= \sum_j \sum_k j\, k\, p_{\mathsf{X}}(j)\, p_{\mathsf{Y}}(k) \\
&= \sum_j j\, p_{\mathsf{X}}(j) \sum_k k\, p_{\mathsf{Y}}(k) \\
&= \left(\sum_j j\, p_{\mathsf{X}}(j)\right) \left(\sum_k k\, p_{\mathsf{Y}}(k)\right) \\
&= \mathsf{E}(\mathsf{X})\ \mathsf{E}(\mathsf{Y})
\end{aligned}
$$

$$\mathsf{COV}(\mathsf{X},\mathsf{Y}) = ??$$

Thus independence implies uncorrelatedness. Does uncorrelatedness imply independence?

Covariance of two Independent Random Variables

Let X and Y be two independent random variables Want to know $\mathsf{E}(\mathsf{XY})$.

$$
\begin{aligned}
\mathsf{E}(\mathsf{XY}) &= \sum_j \sum_k j\, k\, p_{\mathsf{X},\mathsf{Y}}(j,k) \\
&= \sum_j \sum_k j\, k\, p_{\mathsf{X}}(j)\, p_{\mathsf{Y}}(k) \\
&= \sum_j j\, p_{\mathsf{X}}(j) \sum_k k\, p_{\mathsf{Y}}(k) \\
&= \left(\sum_j j\, p_{\mathsf{X}}(j)\right)\left(\sum_k k\, p_{\mathsf{Y}}(k)\right) \\
&= \mathsf{E}(\mathsf{X})\ \mathsf{E}(\mathsf{Y})
\end{aligned}
$$

$$\mathsf{COV}(\mathsf{X},\mathsf{Y}) = ??$$

Thus independence implies uncorrelatedness. Does uncorrelatedness imply independence?

Covariance of two Independent Random Variables

Let X and Y be two independent random variables Want to know $\mathsf{E}(\mathsf{XY})$.

$$
\begin{aligned}
\mathsf{E}(\mathsf{XY}) &= \sum_j \sum_k j\,k\,p_{\mathsf{X},\mathsf{Y}}(j,k) \\
&= \sum_j \sum_k j\,k\,p_{\mathsf{X}}(j)\,p_{\mathsf{Y}}(k) \\
&= \sum_j j\,p_{\mathsf{X}}(j) \sum_k k\,p_{\mathsf{Y}}(k) \\
&= \left(\sum_j j\,p_{\mathsf{X}}(j)\right)\left(\sum_k k\,p_{\mathsf{Y}}(k)\right) \\
&= \mathsf{E}(\mathsf{X})\ \mathsf{E}(\mathsf{Y})
\end{aligned}
$$

$$\mathsf{COV}(\mathsf{X},\mathsf{Y}) = ??$$

Thus independence implies uncorrelatedness. Does uncorrelatedness imply independence?

Covariance of two Independent Random Variables

Let X and Y be two independent random variables Want to know $\mathsf{E}(\mathsf{XY})$.

$$
\begin{aligned}
\mathsf{E}(\mathsf{XY}) &= \sum_j \sum_k j\, k\, p_{\mathsf{X},\mathsf{Y}}(j,k) \\
&= \sum_j \sum_k j\, k\, p_{\mathsf{X}}(j)\, p_{\mathsf{Y}}(k) \\
&= \sum_j j\, p_{\mathsf{X}}(j) \sum_k k\, p_{\mathsf{Y}}(k) \\
&= \left(\sum_j j\, p_{\mathsf{X}}(j)\right)\left(\sum_k k\, p_{\mathsf{Y}}(k)\right) \\
&= \mathsf{E}(\mathsf{X})\ \mathsf{E}(\mathsf{Y})
\end{aligned}
$$

$$
\mathsf{COV}(\mathsf{X},\mathsf{Y}) = ??
$$

Thus independence implies uncorrelatedness. Does uncorrelatedness imply independence?

Covariance of two Independent Random Variables

Let X and Y be two independent random variables Want to know $\mathsf{E}(\mathsf{XY})$.

$$
\begin{aligned}
\mathsf{E}(\mathsf{XY}) &= \sum_j \sum_k j\, k\, p_{\mathsf{X},\mathsf{Y}}(j,k) \\
&= \sum_j \sum_k j\, k\, p_{\mathsf{X}}(j)\, p_{\mathsf{Y}}(k) \\
&= \sum_j j\, p_{\mathsf{X}}(j) \sum_k k\, p_{\mathsf{Y}}(k) \\
&= \left(\sum_j j\, p_{\mathsf{X}}(j)\right)\left(\sum_k k\, p_{\mathsf{Y}}(k)\right) \\
&= \mathsf{E}(\mathsf{X})\ \mathsf{E}(\mathsf{Y})
\end{aligned}
$$

$$
\mathsf{COV}(\mathsf{X},\mathsf{Y}) = ??
$$

Thus independence implies uncorrelatedness. Does uncorrelatedness imply independence?

Covariance Identities

- $\mathsf{COV}(\mathsf{X}+\mathsf{Y},\mathsf{Z}) = \mathsf{COV}(\mathsf{X},\mathsf{Y}) + \mathsf{COV}(\mathsf{X},\mathsf{Z})$
- $\mathsf{COV}(\mathsf{X}_1+\mathsf{X}_2,\mathsf{Y}_1+\mathsf{Y}_2) =$
  $\mathsf{COV}(\mathsf{X}_1,\mathsf{Y}_1) + \mathsf{COV}(\mathsf{X}_1,\mathsf{Y}_2) + \mathsf{COV}(\mathsf{X}_2,\mathsf{Y}_1) + \mathsf{COV}(\mathsf{X}_2,\mathsf{Y}_2)$
- More generally,

$$\mathsf{COV}\left(\sum_{i=1}^{m}\mathsf{X}_i, \sum_{j=1}^{n}\mathsf{Y}_j\right) = \sum_{i,j}\mathsf{COV}(\mathsf{X}_i,\mathsf{Y}_j)$$

- Variance of the sum of random variables

$$\mathsf{VAR}(\mathsf{X}_1+\mathsf{X}_2) = \mathsf{VAR}(\mathsf{X}_1) + \mathsf{VAR}(\mathsf{X}_2) + \mathsf{COV}(\mathsf{X}_1,\mathsf{X}_2)$$

- More generally

$$\mathsf{VAR}\left(\sum_{i=1}^{m}\mathsf{X}_i\right) \quad = \quad ???$$

Covariance Identities

- Variance of the sum of arbitrary random variables

$$\mathsf{VAR}\left(\sum_{i=1}^{m} \mathsf{X}_i\right) = \sum_{i=1}^{m} \mathsf{VAR}(\mathsf{X}_i) + \sum_{\substack{j=1 \\ j \neq i}}^{m} \mathsf{COV}(\mathsf{X}_i, \mathsf{X}_j)$$

- Correlation coefficient of two random variables

$$\mathsf{Corr}(\mathsf{X}, \mathsf{Y}) = \frac{\mathsf{COV}(\mathsf{X}, \mathsf{Y})}{\sqrt{\mathsf{VAR}(\mathsf{X})\ \mathsf{VAR}(\mathsf{Y})}}$$

- Can show that $|\mathsf{Corr}(\mathsf{X}, \mathsf{Y})| \leq 1$
- Notation: $\rho_{\mathsf{XY}} = \mathsf{Corr}(\mathsf{X}, \mathsf{Y})$.

Conditional Distributions

- Conditional distributions

$$
\begin{aligned}
F_{\mathsf{Y}}(y|\mathsf{X} \leq x) &= \frac{\Pr(\mathsf{Y} \leq y, \mathsf{X} \leq x)}{\Pr(\mathsf{X} \leq x)} = \frac{F_{\mathsf{XY}}(x,y)}{F_{\mathsf{X}}(x)} \\
f_{\mathsf{Y}}(y|\mathsf{X} \leq x) &= \;= \frac{\partial F_{\mathsf{XY}}(x,y)/\partial y}{F_{\mathsf{X}}(x)} \\
F_{\mathsf{Y}}(y|x_1 \leq \mathsf{X} \leq x_2) &= \frac{\Pr(\mathsf{Y} \leq y, x_1 < \mathsf{X} \leq x_2)}{\Pr(x_1 \leq \mathsf{X} \leq x_2)} = \frac{F_{\mathsf{XY}}(x_2,y) - F_{\mathsf{XY}}(x_1,y)}{F_{\mathsf{X}}(x_2) - F_{\mathsf{X}}(x_1)} \\
f_{\mathsf{Y}}(y|\mathsf{X} = x) &= \frac{f_{\mathsf{XY}}(x,y)}{f_{\mathsf{X}}(x)}
\end{aligned}
$$

- Total probability

$$f_{\mathsf{X}}(x) = \int_{-\infty}^{\infty} f_{\mathsf{X}}(x|y)\, f_{\mathsf{Y}}(y)\, dy$$

- Bayes theorem

$$f_{\mathsf{Y}}(y|x) = \frac{f_{\mathsf{X}}(x|y)\, f_{\mathsf{Y}}(y)}{\int_{-\infty}^{\infty} f_{\mathsf{X}}(x|y)\, f_{\mathsf{Y}}(y) dy}$$

Conditional Distributions

- Conditional distributions

$$F_Y(y|X \le x) = \frac{\Pr(Y \le y, X \le x)}{\Pr(X \le x)} = \frac{F_{XY}(x,y)}{F_X(x)}$$

$$f_Y(y|X \le x) = \; = \frac{\partial F_{XY}(x,y)/\partial y}{F_X(x)}$$

$$F_Y(y|x_1 \le X \le x_2) = \frac{\Pr(Y \le y, x_1 < X \le x_2)}{\Pr(x_1 \le X \le x_2)} = \frac{F_{XY}(x_2,y) - F_{XY}(x_1,y)}{F_X(x_2) - F_X(x_1)}$$

$$f_Y(y|X = x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

- Total probability

$$f_X(x) = \int_{-\infty}^{\infty} f_X(x|y)\, f_Y(y)\, dy$$

- Bayes theorem

$$f_Y(y|x) = \frac{f_X(x|y)\, f_Y(y)}{\int_{-\infty}^{\infty} f_X(x|y)\, f_Y(y) dy}$$

Conditional Distributions

- Conditional distributions

$$F_Y(y|X \le x) = \frac{\Pr(Y \le y, X \le x)}{\Pr(X \le x)} = \frac{F_{XY}(x,y)}{F_X(x)}$$

$$f_Y(y|X \le x) = \; = \frac{\partial F_{XY}(x,y)/\partial y}{F_X(x)}$$

$$F_Y(y|x_1 \le X \le x_2) = \frac{\Pr(Y \le y, x_1 < X \le x_2)}{\Pr(x_1 \le X \le x_2)} = \frac{F_{XY}(x_2,y) - F_{XY}(x_1,y)}{F_X(x_2) - F_X(x_1)}$$

$$f_Y(y|X = x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

- Total probability

$$f_X(x) = \int_{-\infty}^{\infty} f_X(x|y)\, f_Y(y)\, dy$$

- Bayes theorem

$$f_Y(y|x) = \frac{f_X(x|y)\, f_Y(y)}{\int_{-\infty}^{\infty} f_X(x|y)\, f_Y(y) dy}$$

Conditional Distributions

- Conditional distributions

$$
\begin{aligned}
F_Y(y|X \le x) &= \frac{\Pr(Y \le y, X \le x)}{\Pr(X \le x)} = \frac{F_{XY}(x,y)}{F_X(x)} \\
f_Y(y|X \le x) &= \; = \frac{\partial F_{XY}(x,y)/\partial y}{F_X(x)} \\
F_Y(y|x_1 \le X \le x_2) &= \frac{\Pr(Y \le y, x_1 < X \le x_2)}{\Pr(x_1 \le X \le x_2)} = \frac{F_{XY}(x_2,y) - F_{XY}(x_1,y)}{F_X(x_2) - F_X(x_1)} \\
f_Y(y|X = x) &= \frac{f_{XY}(x,y)}{f_X(x)}
\end{aligned}
$$

- Total probability

$$f_X(x) = \int_{-\infty}^{\infty} f_X(x|y)\, f_Y(y)\, dy$$

- Bayes theorem

$$f_Y(y|x) = \frac{f_X(x|y)\, f_Y(y)}{\int_{-\infty}^{\infty} f_X(x|y)\, f_Y(y) dy}$$

Conditional Distributions

- Conditional distributions

$$F_{\mathsf{Y}}(y|\mathsf{X} \leq x) = \frac{\Pr(\mathsf{Y} \leq y, \mathsf{X} \leq x)}{\Pr(\mathsf{X} \leq x)} = \frac{F_{\mathsf{XY}}(x,y)}{F_{\mathsf{X}}(x)}$$

$$f_{\mathsf{Y}}(y|\mathsf{X} \leq x) = = \frac{\partial F_{\mathsf{XY}}(x,y)/\partial y}{F_{\mathsf{X}}(x)}$$

$$F_{\mathsf{Y}}(y|x_1 \leq \mathsf{X} \leq x_2) = \frac{\Pr(\mathsf{Y} \leq y, x_1 < \mathsf{X} \leq x_2)}{\Pr(x_1 \leq \mathsf{X} \leq x_2)} = \frac{F_{\mathsf{XY}}(x_2,y) - F_{\mathsf{XY}}(x_1,y)}{F_{\mathsf{X}}(x_2) - F_{\mathsf{X}}(x_1)}$$

$$f_{\mathsf{Y}}(y|\mathsf{X} = x) = \frac{f_{\mathsf{XY}}(x,y)}{f_{\mathsf{X}}(x)}$$

- Total probability

$$f_{\mathsf{X}}(x) = \int_{-\infty}^{\infty} f_{\mathsf{X}}(x|y)\, f_{\mathsf{Y}}(y)\, dy$$

- Bayes theorem

$$f_{\mathsf{Y}}(y|x) = \frac{f_{\mathsf{X}}(x|y)\, f_{\mathsf{Y}}(y)}{\int_{-\infty}^{\infty} f_{\mathsf{X}}(x|y)\, f_{\mathsf{Y}}(y) dy}$$

Conditional Expectation

- Conditional expectation of X given $\mathsf{Y} = y$ is

$$\mathsf{E}(\mathsf{X}|\mathsf{Y} = y) = \int_{-\infty}^{\infty} x\, f_{\mathsf{X}}(x|\mathsf{Y} = y)\, dx$$

- Observe that this is a function of y and is a random variable—depends on the value of Y from the experiment.
- Can define its moments. Leading us to chain rule of expectations:

$$\begin{aligned}
\mathsf{E}(\mathsf{E}(\mathsf{X}|\mathsf{Y})) &= \int_{-\infty}^{\infty} \mathsf{E}(\mathsf{X}|\mathsf{Y})\, f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x\, f_{\mathsf{X}|\mathsf{Y}}(x|y)\, dx \right) f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{X}|\mathsf{Y}}(x, y) f_{\mathsf{Y}}(y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{XY}}(x, y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x f_{\mathsf{X}}(x)\, dx = \mathsf{E}(\mathsf{X})
\end{aligned}$$

Conditional Expectation

- Conditional expectation of X given $\mathsf{Y} = y$ is

$$\mathsf{E}(\mathsf{X}|\mathsf{Y}=y) = \int_{-\infty}^{\infty} x\, f_{\mathsf{X}}(x|\mathsf{Y}=y)\, dx$$

- Observe that this is a function of y and is a random variable—depends on the value of Y from the experiment.
- Can define its moments. Leading us to chain rule of expectations:

$$\begin{aligned}
\mathsf{E}(\mathsf{E}(\mathsf{X}|\mathsf{Y})) &= \int_{-\infty}^{\infty} \mathsf{E}(\mathsf{X}|\mathsf{Y})\, f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x\, f_{\mathsf{X}|\mathsf{Y}}(x|y)\, dx \right) f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{X}|\mathsf{Y}}(x,y) f_{\mathsf{Y}}(y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{XY}}(x,y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x f_{\mathsf{X}}(x)\, dx = \mathsf{E}(\mathsf{X})
\end{aligned}$$

Conditional Expectation

- Conditional expectation of X given $\mathsf{Y} = y$ is

$$\mathsf{E}(\mathsf{X}|\mathsf{Y} = y) = \int_{-\infty}^{\infty} x\, f_{\mathsf{X}}(x|\mathsf{Y} = y)\, dx$$

- Observe that this is a function of y and is a random variable—depends on the value of Y from the experiment.
- Can define its moments. Leading us to chain rule of expectations:

$$\begin{aligned}
\mathsf{E}(\mathsf{E}(\mathsf{X}|\mathsf{Y})) &= \int_{-\infty}^{\infty} \mathsf{E}(\mathsf{X}|\mathsf{Y})\, f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x\, f_{\mathsf{X}|\mathsf{Y}}(x|y)\, dx \right) f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{X}|\mathsf{Y}}(x, y) f_{\mathsf{Y}}(y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{XY}}(x, y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x f_{\mathsf{X}}(x)\, dx = \mathsf{E}(\mathsf{X})
\end{aligned}$$

Conditional Expectation

- Conditional expectation of X given $\mathsf{Y} = y$ is

$$\mathsf{E}(\mathsf{X}|\mathsf{Y} = y) = \int_{-\infty}^{\infty} x\, f_{\mathsf{X}}(x|\mathsf{Y} = y)\, dx$$

- Observe that this is a function of y and is a random variable—depends on the value of Y from the experiment.
- Can define its moments. Leading us to chain rule of expectations:

$$\begin{aligned}
\mathsf{E}(\mathsf{E}(\mathsf{X}|\mathsf{Y})) &= \int_{-\infty}^{\infty} \mathsf{E}(\mathsf{X}|\mathsf{Y})\, f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x\, f_{\mathsf{X}|\mathsf{Y}}(x|y)\, dx \right) f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{X}|\mathsf{Y}}(x,y) f_{\mathsf{Y}}(y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{XY}}(x,y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x f_{\mathsf{X}}(x)\, dx = \mathsf{E}(\mathsf{X})
\end{aligned}$$

Conditional Expectation

- Conditional expectation of X given $\mathsf{Y} = y$ is

$$\mathsf{E}(\mathsf{X}|\mathsf{Y} = y) = \int_{-\infty}^{\infty} x\, f_{\mathsf{X}}(x|\mathsf{Y} = y)\, dx$$

- Observe that this is a function of y and is a random variable—depends on the value of Y from the experiment.
- Can define its moments. Leading us to chain rule of expectations:

$$\begin{aligned}
\mathsf{E}(\mathsf{E}(\mathsf{X}|\mathsf{Y})) &= \int_{-\infty}^{\infty} \mathsf{E}(\mathsf{X}|\mathsf{Y})\, f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x\, f_{\mathsf{X}|\mathsf{Y}}(x|y)\, dx \right) f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{X}|\mathsf{Y}}(x, y) f_{\mathsf{Y}}(y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{XY}}(x, y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x f_{\mathsf{X}}(x)\, dx = \mathsf{E}(\mathsf{X})
\end{aligned}$$

Conditional Expectation

- Conditional expectation of X given $\mathsf{Y} = y$ is

$$\mathsf{E}(\mathsf{X}|\mathsf{Y}=y) = \int_{-\infty}^{\infty} x\, f_{\mathsf{X}}(x|\mathsf{Y}=y)\, dx$$

- Observe that this is a function of y and is a random variable—depends on the value of Y from the experiment.
- Can define its moments. Leading us to chain rule of expectations:

$$\begin{aligned}
\mathsf{E}(\mathsf{E}(\mathsf{X}|\mathsf{Y})) &= \int_{-\infty}^{\infty} \mathsf{E}(\mathsf{X}|\mathsf{Y})\, f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x\, f_{\mathsf{X}|\mathsf{Y}}(x|y)\, dx \right) f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{X}|\mathsf{Y}}(x,y) f_{\mathsf{Y}}(y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{XY}}(x,y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x f_{\mathsf{X}}(x)\, dx \;=\; \mathsf{E}(\mathsf{X})
\end{aligned}$$

Conditional Expectation

- Conditional expectation of X given $\mathsf{Y} = y$ is

$$\mathsf{E}(\mathsf{X}|\mathsf{Y} = y) = \int_{-\infty}^{\infty} x\, f_{\mathsf{X}}(x|\mathsf{Y} = y)\, dx$$

- Observe that this is a function of y and is a random variable—depends on the value of Y from the experiment.
- Can define its moments. Leading us to chain rule of expectations:

$$\begin{aligned}
\mathsf{E}(\mathsf{E}(\mathsf{X}|\mathsf{Y})) &= \int_{-\infty}^{\infty} \mathsf{E}(\mathsf{X}|\mathsf{Y})\, f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x\, f_{\mathsf{X}|\mathsf{Y}}(x|y)\, dx \right) f_{\mathsf{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{X}|\mathsf{Y}}(x, y) f_{\mathsf{Y}}(y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\mathsf{XY}}(x, y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x f_{\mathsf{X}}(x)\, dx = \mathsf{E}(\mathsf{X})
\end{aligned}$$