Lecture Notes 4
Expectation

- Definition and Properties
- Mean and Variance
- Markov and Chebyshev Inequalities
- Covariance and Correlation
- Conditional Expectation

Expectation

- Let $X \in \mathcal{X}$ be a discrete r.v. with pmf $p_X(x)$ and let $g(x)$ be a function of x. The *expectation* (or *expected value* or *mean*) of $g(X)$ can be defined as

$$\mathrm{E}(g(X)) = \sum_{x \in \mathcal{X}} g(x) p_X(x)$$

- For a continuous r.v. $X \sim f_X(x)$, the expected value of $g(X)$ can be defined as

$$\mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

- Properties of expectation:
  - If c is a constant, then $\mathrm{E}(c) = c$
  - Expectation is *linear*, i.e., for any constants a, b

$$\mathrm{E}[a g_1(X) + b g_2(X)] = a\, \mathrm{E}(g_1(X)) + b\, \mathrm{E}(g_2(X))$$

◦ Fundamental Theorem of Expectation: If $Y = g(X) \sim p_Y(y)$, then

$$\mathrm{E}(Y) = \sum_{y \in \mathcal{Y}} y p_Y(y) = \sum_{x \in \mathcal{X}} g(x) p_X(x) = \mathrm{E}(g(X))$$

The corresponding formula for $f_Y(y)$ uses integrals instead of sums:

$$\mathrm{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)\, dy$$

Proof: We prove the theorem for discrete r.v.s. Consider

$$\begin{aligned}\mathrm{E}(Y) &= \sum_y y p_Y(y) \\ &= \sum_y y \sum_{\{x:\, g(x)=y\}} p_X(x) \\ &= \sum_y \sum_{\{x:\, g(x)=y\}} y p_X(x) = \sum_y \sum_{\{x:\, g(x)=y\}} g(x) p_X(x) = \sum_x g(x) p_X(x)\end{aligned}$$

Thus $\mathrm{E}(Y) = \mathrm{E}(g(X))$ can be found using either $f_X(x)$ or $f_Y(y)$

It is often much easier to use $f_X(x)$ than to first find $f_Y(y)$ and then find $\mathrm{E}(Y)$

- Remark: We know that a r.v. is completely specified by its cdf (pdf, pmf), so why do we need expectation?
  - Expectation provides a *summary* or an *estimate* of the r.v. — a single number — instead of specifying the entire distribution.
  - It is far easier to estimate the expectation of a r.v. from data than to estimate its distribution
  - Expectation can be used to bound or estimate probabilities of interesting events (as we shall see)
  - We will later see that expections predict the numerical average of a sequence of measurements. For example, if each minute a sensor measures a random variable, and if the sequence of random variables is independent and has the same pdf, then in some sense the sample mean formed by adding up all the measurements and dividing by the number of measurements will be close to the common expectations of the random variable. (Law of large numbers)
  - Many problems in signal processing, such as estimating and predicting future values based on past observations, have solutions in terms of expectations.

Mean and Variance

- The *first moment* (or *mean*) of $X \sim f_X(x)$ is the expectation

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

- The *second moment* (or *mean square* or *average power*) of X is

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx$$

- The *variance* of X is

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] \\ &= \mathrm{E}\left[X^2 - 2X\,\mathrm{E}(X) + (\mathrm{E}(X))^2\right] \\ &= \mathrm{E}(X^2) - 2(\mathrm{E}(X))^2 + (\mathrm{E}(X))^2 \\ &= \mathrm{E}(X^2) - (\mathrm{E}(X))^2\end{aligned}$$

- The *standard deviation* of X is defined as $\sigma_X = \sqrt{\mathrm{Var}(X)}$, i.e., $\mathrm{Var}(X) = \sigma^2$

Mean and Variance for Famous R.V.s

Random Variable	Mean	Variance
$\mathrm{Bern}(p)$	p	$p(1-p)$
$\mathrm{Geom}(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$\mathrm{Binom}(n,p)$	np	$np(1-p)$
$\mathrm{Poisson}(\lambda)$	λ	λ
$\mathrm{U}[\,a,b\,]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\mathrm{Exp}(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\mathcal{N}(\mu,\sigma^2)$	μ	σ^2

Expectation Can Be Infinite or Might Not Exist

- Expectation can be infinite. For example
$$f_X(x) = \begin{cases} 1/x^2 & 1 \le x < \infty \\ 0 & \text{otherwise} \end{cases} \Rightarrow \mathrm{E}(X) = \int_1^\infty x/x^2\, dx = \infty$$
- Expectation might not exist. To find conditions for expectation to exist, consider
$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx = -\int_{-\infty}^{0} |x| f_X(x)\, dx + \int_0^\infty |x| f_X(x)\, dx\,,$$
so either $\int_{-\infty}^{0} |x| f_X(x)\, dx$ or $\int_0^\infty |x| f_X(x)\, dx$ must be finite
- Example: the *standard Cauchy* r.v. has the pdf
$$f(x) = \frac{1}{\pi(1+x^2)}$$
Since both $\int_{-\infty}^{0} |x| f(x)\, dx$ and $\int_0^\infty |x| f(x)\, dx$ are infinite, its mean does not exist! (The second moment of the Cauchy is $\mathrm{E}(X^2) = \infty$, so it exists)

Bounding Probability Using Expectation

- In many cases we do not know the distribution of a r.v. X but want to find the probability of an event such as $\{X > a\}$ or $\{|X - \mathrm{E}(X)| > a\}$
- The Markov and Chebyshev inequalities give upper bounds on the probabilities of such events in terms of the mean and variance of the random variable
- Example: Let $X \ge 0$ represent the age of a person in the Bay Area. If we know that $\mathrm{E}(X) = 35$ years, what fraction of the population is ≥ 70 years old?

  Clearly we cannot answer this question knowing only the mean, but we can say that $\mathrm{P}\{X \ge 70\} \le 0.5$, since otherwise the mean would be larger than 35
- This is an application of the *Markov inequality*

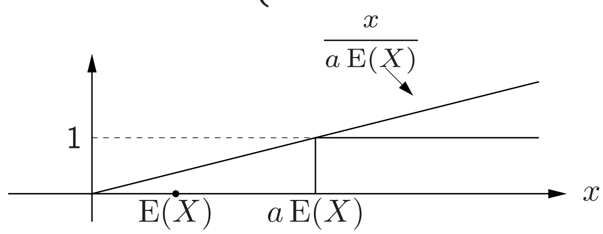
Markov Inequality

- For any r.v. $X \geq 0$ with finite mean $\mathrm{E}(X)$ and any $a > 0$,

$$\mathrm{P}\{X \geq a\,\mathrm{E}(X)\} \leq \frac{1}{a}$$

Proof: Define the *indicator function* of the set $A = \{x \geq a\,\mathrm{E}(X)\}$:

$$\mathrm{I}_A(x) = \begin{cases} 1 & x \geq a\,\mathrm{E}(X) \\ 0 & \text{otherwise} \end{cases}$$



Clearly, $\mathrm{I}_A(x) \leq \dfrac{x}{a\,\mathrm{E}(X)}$

Since $\mathrm{E}(\mathrm{I}_A) = \mathrm{P}(A) = \mathrm{P}\{X \geq a\,\mathrm{E}(X)\}$, taking the expectations of both sides we obtain the Markov Inequality

- The Markov inequality can be *very* loose. If $X \sim \mathrm{Exp}(1)$, then

$$\mathrm{P}\{X \geq 10\} = e^{-10} \approx 4.54 \times 10^{-5}$$

The Markov inequality gives

$$\mathrm{P}\{X \geq 10\} \leq \tfrac{1}{10}\,,$$

which is very pessimistic

- But, it is the *tightest* possible inequality on $\mathrm{P}\{X \geq a\,\mathrm{E}(X)\}$ when we are given only the mean of X

To show this, note that the inequality is tight for the following r.v.:

$$X = \begin{cases} a\,\mathrm{E}(X) & \text{with probability } 1/a \\ 0 & \text{with probability } 1 - 1/a \end{cases}$$

Chebyshev Inequality

- Let X be a device parameter in an integrated circuit (IC) with known mean and variance. The IC is out-of-spec if X is more than, say, $3\sigma_X$ away from its mean. We wish to find the fraction of out-of-spec ICs, namely, $\mathrm{P}\{|X - \mathrm{E}(X)| \geq 3\sigma_X\}$

  The *Chebyshev inequality* gives us an upper bound on this fraction in terms the mean and variance of X

- Let X be a r.v. with known $\mathrm{E}(X)$ and $\mathrm{Var}(X) = \sigma_X^2$. The Chebyshev inequality states that for every $a > 1$,

$$\mathrm{P}\{|X - \mathrm{E}(X)| \geq a\sigma_X\} \leq \frac{1}{a^2}$$

  Proof: We use the Markov inequality. Define the r.v. $Y = (X - \mathrm{E}(X))^2 \geq 0$. Since $\mathrm{E}(Y) = \sigma_X^2$, the Markov inequality gives

$$\mathrm{P}\{Y \geq a^2\sigma_X^2\} \leq \frac{1}{a^2}$$

  But $\{|X - \mathrm{E}(X)| \geq a\sigma_X\}$ occurs iff $\{Y \geq a^2\sigma_X^2\}$. Thus

$$\mathrm{P}\{|X - \mathrm{E}(X)| \geq a\sigma_X\} \leq \frac{1}{a^2}$$

- The Chebyshev inequality can be very loose. Let $X \sim \mathcal{N}(0,1)$. Using the Chebyshev inequality we obtain

$$\mathrm{P}\{|X| \geq 3\} \leq \tfrac{1}{9},$$

  which is very pessimistic compared to the actual value $2Q(3) \approx 2 \times 10^{-3}$

- But, it is the tightest inequality on $\mathrm{P}\{|X - \mathrm{E}(X)| \geq a\sigma_X\}$ given knowledge only of the mean and variance of X. To show this, note that equality is achieved for the random variable

$$X = \begin{cases} \mathrm{E}(X) + a\sigma_X & \text{with probability } 1/2a^2 \\ \mathrm{E}(X) - a\sigma_X & \text{with probability } 1/2a^2 \\ \mathrm{E}(X) & \text{with probability } 1 - 1/a^2 \end{cases}$$

Expectation Involving Two RVs

- Let $(X, Y) \sim f_{X,Y}(x, y)$ and let $g(x, y)$ be a function of x and y. The expectation of $g(X, Y)$ is given by

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

The function $g(X, Y)$ may be X, Y, X^2, $X + Y$, etc.

- The *correlation* of X and Y is defined as $\mathrm{E}(XY)$
  X and Y are said to be *orthogonal* if $\mathrm{E}(XY) = 0$
- The *covariance* of X and Y is defined as

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

X and Y are said to be *uncorrelated* if $\mathrm{Cov}(X, Y) = 0$

- Note that $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$

- Example: Find $\mathrm{E}(X)$, $\mathrm{Var}(X)$, and $\mathrm{Cov}(X, Y)$ for $(X, Y) \sim f(x, y)$ where

$$f(x, y) \; = \begin{cases} 2 & x \ge 0,\, y \ge 0,\, x + y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Solution:

$$\begin{aligned}\mathrm{E}(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x, y)\, dx\, dy \\ &= \int_0^1 \int_0^{1-x} 2x\, dy\, dx \;=\; 2\int_0^1 (1 - x)x\, dx \;=\; 2(\tfrac{1}{2} - \tfrac{1}{3}) \;=\; \tfrac{1}{3}\end{aligned}$$

Since $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$, we need to find the second moment:

$$\mathrm{E}(X^2) \;=\; 2\int_0^1 (1 - x)x^2\, dx \;=\; 2(\tfrac{1}{3} - \tfrac{1}{4}) \;=\; \tfrac{1}{6}\,,$$

hence

$$\mathrm{Var}(X) \;=\; \tfrac{1}{6} - (\tfrac{1}{3})^2 \;=\; \tfrac{1}{6} - \tfrac{1}{9} \;=\; \tfrac{1}{18}$$

By symmetry, $\mathrm{E}(Y) = \mathrm{E}(X) = \frac{1}{3}$. Thus the covariance of X and Y is

$$\begin{aligned}\mathrm{Cov}(X,Y) &= 2\int_0^1 \int_0^{1-x} xy\,dy\,dx - \mathrm{E}(X)\,\mathrm{E}(Y) \\ &= \int_0^1 x(1-x)^2\,dx - \tfrac{1}{9} \;=\; \tfrac{1}{12} - \tfrac{1}{9} = -\tfrac{1}{36}\end{aligned}$$

Uncorrelation vs. Independence

- Recall that X and Y are *uncorrelated* if $\mathrm{Cov}(X,Y) = 0$
- If X and Y are independent then they are uncorrelated, since

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

  Therefore $\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) = 0$

- X and Y uncorrelated does *not* necessarily imply that they are independent

- Example: Let $X, Y \in \{-2, -1, +1, +2\}$ such that

$$p(x, y) = \begin{cases} \frac{2}{5} & (x, y) = (+1, +1), (-1, -1) \\ \frac{1}{10} & (x, y) = (+2, -2), (-2, +2) \\ 0 & \text{otherwise} \end{cases}$$

Are X and Y independent? Are they uncorrelated?

- Solution: Clearly X and Y are not independent

Let's check their covariance:

$$\mathrm{E}(X) = \tfrac{2}{5} - \tfrac{2}{5} - \tfrac{2}{10} + \tfrac{2}{10} = 0$$
$$\mathrm{E}(Y) = 0 \quad \text{(by symmetry)}$$
$$\mathrm{E}(XY) = \tfrac{2}{5} + \tfrac{2}{5} - \tfrac{4}{10} - \tfrac{4}{10} = 0$$

Thus, $\mathrm{Cov}(X, Y) = 0$, and X and Y are uncorrelated!

Correlation Coefficient

- The *correlation coefficient* of X and Y is defined as

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

- Fact: $|\rho_{X,Y}| \le 1$ (a form of the *Cauchy-Schwartz inequality*) To show this consider

$$\mathrm{E}\left[\left(\frac{X - \mathrm{E}(X)}{\sigma_X} \pm \frac{Y - \mathrm{E}(Y)}{\sigma_Y}\right)^2\right] \ge 0$$

$$\frac{\mathrm{E}\left[(X - \mathrm{E}(X))^2\right]}{\sigma_X^2} + \frac{\mathrm{E}\left[(Y - \mathrm{E}(Y))^2\right]}{\sigma_Y^2} \pm 2\frac{\mathrm{E}\left[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right]}{\sigma_X \sigma_Y} \ge 0$$

$$1 + 1 \pm 2\rho_{X,Y} \ge 0 \;\Rightarrow\; -2 \le 2\rho_{X,Y} \le 2 \;\Rightarrow\; |\rho_{X,Y}| \le 1$$

- From the proof, we see that $\rho_{X,Y} = \pm 1$ iff $\dfrac{X - \mathrm{E}(X)}{\sigma_X} = \pm \dfrac{Y - \mathrm{E}(Y)}{\sigma_Y}$ (equality with probability 1), i.e., iff $X - \mathrm{E}(X)$ is a linear function of $Y - \mathrm{E}(Y)$

- Note: We shall see that $\rho_{X,Y}$ is a measure of how closely $X - \mathrm{E}(X)$ can be approximated or estimated by a linear function of $Y - \mathrm{E}(Y)$

Conditional Expectation

- Conditioning on an event: Let $X \sim p_X(x)$ and A be a nonzero probability event. We can define the conditional pmf of X given $X \in A$ as

$$p_{X|A}(x) = \mathrm{P}\{X = x \,|\, X \in A\} = \frac{\mathrm{P}\{X = x, X \in A\}}{\mathrm{P}\{X \in A\}} = \begin{cases} \frac{p_X(x)}{\mathrm{P}\{X \in A\}} & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

  Note that $p_{X|A}(x)$ is a pmf on X

- Similarly, if $X \sim f_X(x)$,

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{\mathrm{P}\{X \in A\}} & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

  is a pdf on X

- Example: Let $X \sim \mathrm{Exp}(\lambda)$ and $A = \{X > a\}$, for some $a > 0$. The conditional pdf of X given A is $\lambda e^{-\lambda(x-a)}$, for $x - a \geq 0$,and 0 otherwise

- We define the *conditional expectation* of $g(X)$ given $X \in A$ as

$$\mathrm{E}(g(X) \,|\, A) = \begin{cases} \int_{-\infty}^{\infty} g(x) f_{X|A}(x)\, dx & X \,|\, A \text{ continuous} \\ \sum_{x=-\infty}^{\infty} g(x) p_{X|A}(x) & X \,|\, A \text{ discrete} \end{cases}$$

- Example: Find $\mathrm{E}(X \,|\, A)$ and $\mathrm{E}(X^2 \,|\, A)$ for the previous example

- Total Expectation Theorem: Let $X \sim f_X(x)$ and $A_1, A_2, \ldots, A_n$ be disjoint nonzero probability events with $\mathrm{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathrm{P}(A_i) = 1$. Then

$$\mathrm{E}(g(X)) = \sum_{i=1}^{n} \mathrm{P}\{X \in A_i\}\, \mathrm{E}(g(X)|A_i)$$

  Proof: First note that by the law of total probability

$$f_X(x) = \sum_{i=1}^{n} \mathrm{P}(A_i) f_{X|A_i}(x)$$

Therefore

$$\begin{aligned}
\mathrm{E}(g(X)) &= \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \\
&= \int_{-\infty}^{\infty} g(x) \sum_{i=1}^{n} \mathrm{P}(A_i) f_{X|A_i}(x)\, dx \\
&= \sum_{i=1}^{n} \mathrm{P}(A_i) \int_{-\infty}^{\infty} g(x) f_{X|A_i}(x)\, dx = \sum_{i=1}^{n} \mathrm{P}(A_i)\, \mathrm{E}(g(X) \mid A_i)
\end{aligned}$$

Similar proof for discrete/pmf case.

This result is useful in computing expectation by *divide-and-conquer*

- Example: Mean and variance of piecewise uniform pdf. Let X be a continuous r.v. with the piecewise uniform pdf

$$f_X(x) = \begin{cases} 1/3 & \text{if } 0 \le x \le 1 \\ 2/3 & \text{if } 1 < x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the mean and variance of X

Solution: The events $A_1 = \{X \in [0,1]\}$ and $A_2 = \{X \in (1,2]\}$ are disjoint and the sum of their probabilities is 1. The mean and second moment of X can be expressed as

$$\mathrm{E}(X) = \sum_{i=1}^{2} \mathrm{P}\{X \in A_i\}\, \mathrm{E}(X \mid A_i) = \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{3}{2} = \frac{7}{6}$$

$$\mathrm{E}(X^2) = \sum_{i=1}^{2} \mathrm{P}\{X \in A_i\}\, \mathrm{E}(X^2 \mid A_i) = \frac{1}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{7}{3} = \frac{15}{9}$$

Thus

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \frac{11}{36}$$

Conditioning on a RV

- Let $(X,Y) \sim f_{X,Y}(x,y)$. If $f_Y(y) \neq 0$, the *conditional pdf* of X given $Y = y$ is given by

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

- We know that $f_{X|Y}(x|y)$ is a pdf (or $p_{X|Y}(x|y)$ is a pmf) for X (function of y), so we can define the expectation of any function $g(X,Y)$ w.r.t. $f_{X|Y}(x|y)$ (or $p_{X|Y}(x|y)$) as

$$\mathrm{E}(g(X,Y) \,|\, Y = y) = \begin{cases} \int_{-\infty}^{\infty} g(x,y) f_{X|Y}(x|y)\,dx & X \,|\, Y \text{ continuous} \\ \sum_{x=-\infty}^{\infty} g(x,y) p_{X|Y}(x|y) & X \,|\, Y \text{ discrete} \end{cases}$$

- Example: If $g(X,Y) = X$, then the conditional expectation of X given $Y = y$ is

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

- Example: If $g(X,Y) = XY$, then $\mathrm{E}(XY \,|\, Y = y) = y\,\mathrm{E}(X \,|\, Y = y)$

- Example: Let

$$f_{X,Y}(x,y) = \begin{cases} 2 & \text{if } x \geq 0,\ y \geq 0,\ x + y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $\mathrm{E}(X \,|\, Y = y)$ and $\mathrm{E}(XY \,|\, Y = y)$

Solution: We already know that

$$f_{X|Y}(x|y) = \begin{cases} \frac{1}{1-y} & \text{if } x \geq 0,\ y \geq 0,\ x + y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Thus

$$\begin{aligned} \mathrm{E}(X|Y = y) &= \int_0^{1-y} \frac{1}{1-y} x\,dx \\ &= \frac{1-y}{2}, \quad 0 \leq y < 1 \end{aligned}$$

Now to find $\mathrm{E}(XY \,|\, Y = y)$, note that

$$\begin{aligned} \mathrm{E}(XY \,|\, Y = y) &= y\,\mathrm{E}(X \,|\, Y = y) \\ &= \frac{y(1-y)}{2}, \quad 0 \leq y < 1 \end{aligned}$$

Conditional Expectation as a RV

- We define the *conditional expectation* of $g(X, Y)$ given Y as the random variable $\mathrm{E}(g(X, Y) \mid Y)$, which is a function of the random variable Y
- In particular, $\mathrm{E}(X \mid Y)$ is the conditional expectation of X given Y, a r.v. that is a function of Y
- Example (continuation of previous example): Find the pdf of $\mathrm{E}(X \mid Y)$

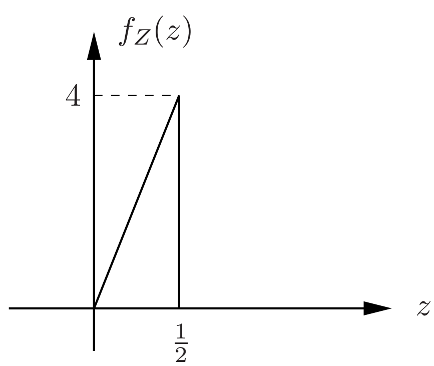
  Solution: The conditional expectation of X given Y is the r.v.

  $$\mathrm{E}(X \mid Y) = \frac{1 - Y}{2} \stackrel{\triangle}{=} Z$$

  The pdf of Z is given by

  $$f_Z(z) = 8z\,, \quad 0 < z \le \tfrac{1}{2}$$

Graph of $f_Z(z)$:



Now let's find the expected value of the r.v. Z

$$\mathrm{E}(Z) = \int_0^{\frac{1}{2}} 8z^2 dz = \tfrac{1}{3} = \mathrm{E}(X)$$

i.e., for this example $\mathrm{E}[\mathrm{E}(X \mid Y)] = \mathrm{E}(X)$. This is in fact true for any X and Y

Iterated Expectation

- In general we can find $\mathrm{E}(g(X,Y))$ using *iterated expectation* as

$$\mathrm{E}(g(X,Y)) = \mathrm{E}_Y\left[\mathrm{E}_X(g(X,Y)\,|\,Y)\right],$$

where E_X means expectation w.r.t. $f_{X|Y}(x|y)$ and E_Y means expectation w.r.t. $f_Y(y)$. To show this, consider

$$\begin{aligned}\mathrm{E}_Y\left[\mathrm{E}_X(g(X,Y)\,|\,Y)\right] &= \int_{-\infty}^{\infty} \mathrm{E}_X(g(X,Y)\,|\,Y=y) f_Y(y)\,dy \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} g(x,y) f_{X|Y}(x|y)\,dx\right) f_Y(y)\,dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X|Y}(x|y) f_Y(y)\,dx\,dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy = \mathrm{E}(g(X,Y))\end{aligned}$$

- This result can be very useful in computing expectation

- Example: A coin has random bias $P \in [0,1]$ with pdf $f_P(p) = 2(1-p)$. The coin is flipped n times. Let N be the number of heads. Find $\mathrm{E}(N)$

Solution: Of course, we could first find the pmf of N, then find its expectation. Using iterated expectation we can find $\mathrm{E}(N)$ more easily

$$\begin{aligned}\mathrm{E}(N) &= \mathrm{E}_P[\mathrm{E}_N(N\,|\,P)] \\ &= \mathrm{E}_P(nP) \\ &= n\int_0^1 2(1-p)p\,dp = \tfrac{1}{3}n\end{aligned}$$

- Example: Let $\mathrm{E}(X\,|\,Y) = Y^2$ and $Y \sim \mathrm{U}[0,1]$. Find $\mathrm{E}(X)$

Solution: We cannot first find the pdf of X, since we do not know $f_{X|Y}(x|y)$, but using iterated expectation we can easily find

$$\mathrm{E}(X) = \mathrm{E}_Y(\mathrm{E}_X(X\,|\,Y)) = \int_0^1 y^2\,dy = \tfrac{1}{3}$$

Conditional Variance

- Let X and Y be two r.v.s. We define the *conditional variance* of X given $Y = y$ to be the variance of X using $f_{X|Y}(x|y)$, i.e.,

$$\begin{aligned}\mathrm{Var}(X \,|\, Y = y) &= \mathrm{E}\left[(X - \mathrm{E}(X \,|\, Y = y))^2 \,|\, Y = y\right] \\ &= \mathrm{E}(X^2 \,|\, Y = y) - \left[\mathrm{E}(X \,|\, Y = y)\right]^2\end{aligned}$$

- The r.v. $\mathrm{Var}(X \,|\, Y)$ is simply a function of Y that takes on the values $\mathrm{Var}(X \,|\, Y = y)$. Its expected value is

$$\mathrm{E}_Y\left[\mathrm{Var}(X \,|\, Y)\right] = \mathrm{E}_Y\left[\mathrm{E}(X^2 \,|\, Y) - (\mathrm{E}(X \,|\, Y))^2\right] = \mathrm{E}(X^2) - \mathrm{E}\left[(\mathrm{E}(X \,|\, Y))^2\right]$$

- Since $\mathrm{E}(X \,|\, Y)$ is a r.v., it has a variance

$$\mathrm{Var}(\mathrm{E}(X \,|\, Y)) = \mathrm{E}_Y\left[\left(\mathrm{E}(X \,|\, Y) - \mathrm{E}[\mathrm{E}(X \,|\, Y)]\right)^2\right] = \mathrm{E}\left[(\mathrm{E}(X \,|\, Y))^2\right] - (\mathrm{E}(X))^2$$

- *Law of Conditional Variances*: Adding the above expressions, we obtain

$$\mathrm{Var}(X) = \mathrm{E}\left(\mathrm{Var}(X \,|\, Y)\right) + \mathrm{Var}\left(\mathrm{E}(X \,|\, Y)\right)$$