INEQUALITIES, DISTRIBUTIONS, CONTINUOUS PROBABILITY, CONDITIONAL EXPECTATION 9

COMPUTER SCIENCE MENTORS 70

November 7 to November 11, 2016

1 Markov and Chebyshev Inequalities

1.1 Introduction

Markov's Inequality
For a non-negative random variable X with expectation $\mathrm{E}(X) = \mu$, and any $\alpha > 0$:

$$\mathrm{P}[X \geq \alpha] \leq \frac{\mathrm{E}(X)}{\alpha}$$

Solution: Proof of Markov's Inequality

$$\begin{aligned}
E(X) &= \sum_a a * Pr[X = a] \\
&\geq \sum a \geq \alpha a * Pr[X = a] \\
&\geq \alpha \sum_{a \geq \alpha} Pr[X = a] \\
&= \alpha Pr[X \geq a]
\end{aligned}$$

Chebyshev's Inequality
For a random variable X with expectation $\mathrm{E}(X) = \mu$, and any $\alpha > 0$:

$$\mathrm{P}[|X - \mu| \geq \alpha] \leq \frac{\mathrm{Var}(X)}{\alpha^2}$$

1.2 Questions

1. Use Markov's to prove Chebyshev's Inequality:

Solution: Define the random variable $Y = (X - \mu)^2$. Note that $\mathrm{E}(Y) = \mathrm{E}((X - \mu)^2) = \mathrm{Var}(X)$ Also, notice that the event that we are interested in, $|X\mu|\alpha$ is exactly the same as the event $Y = (X\mu)^2\alpha^2$. Therefore, $\mathrm{P}[|X\mu| \geq \alpha] = \mathrm{P}[Y \geq \alpha^2]$. Moreover, Y is non-negative, so we can apply Markov's inequality to it to get:

$$\mathrm{P}[Y \geq \alpha^2] \leq \frac{\mathrm{E}(Y)}{\alpha^2} = \frac{\mathrm{Var}(X)}{\alpha^2}$$

2. **Squirrel Standard Deviation**
As we all know, Berkeley squirrels are extremely fat and cute. The average squirrel is 40% body fat. The standard deviation of body fat is 5%. Provide an upper bound on the probability that a randomly trapped squirrel is either too skinny or too fat? A skinny squirrel has less than 27.5% body fat, and a fat squirrel has more than 52.5% body fat?

Solution: We use Chebyshev's inequality. We are looking for the probability we fall within 2.5 standard deviations of the mean. By Chebyshev's inequality, the probability we are within this range is $\frac{1}{2.5^2}, or \frac{4}{25} = 0.16$ If we were to use Markov's inequality, we would probabilities over 1, which yields a non-helpful value.

3. **Bound It**
A random variable X is always strictly larger than -100. You know that $\mathrm{E}(X) = 60$. Give the best upper bound you can on $\mathrm{P}[X \geq 20]$.

Solution: Notice that we do not have the variance of X, so Chebyshev's bound is not applicable here. There is no upper bound on X, so Hoeffdings inequality cannot be used. We know nothing else about its distribution so we cannot evaluate $\mathrm{E}[esX]$ and so Chernoff bounds are not available. Since X is also not a sum of other random variables, other bounds or approximations are not available. This leaves us with just Markov's Inequality. But Markov Bound only applies on a non-negative random variable, whereas X can take on negative values.

This suggests that we want to shift X somehow, so that we can apply Markovs Inequality on it. Define a random variable $Y = X + 100$, which means Y is strictly larger than 0, since X is always strictly larger than 100. Then, $\mathrm{E}(Y) = \mathrm{E}(X+100) = \mathrm{E}(X) + 100 = 60 + 100 = 40$. Finally, the upper bound on X that we want can be calculated via Y, and we can now apply Markov's Inequality on Y since Y is strictly positive.

$\mathrm{P}[X \geq 20] = \mathrm{P}[Y \geq 80] \leq \frac{\mathrm{E}(Y)}{80} = \frac{40}{80} = \frac{1}{2}$

Hence, the best upper bound on $\mathrm{P}[X \geq 20]$ is $\frac{1}{2}$.

4. Give a distribution for a random variable where the expectation is 1,000,000 and the probability that the random variable is zero is 99%.

Solution: X is 100, 000, 000 with probability 0.01, and 0 otherwise.

5. Consider a random variable Y with expectation μ whose maximum value is $\frac{3\mu}{2}$, prove that the probability that Y is 0 is at most $\frac{1}{3}$.

Solution:

$$
\begin{aligned}
\mu &= \sum_a a * \mathrm{P}[Y = a] \\
&= \sum_{a \neq 0} a * \mathrm{P}[Y = a] \\
&\leq \sum_{a \neq 0} \frac{3\mu}{2} * \mathrm{P}[Y = a] \\
&= \frac{3\mu}{2} * \sum_{a \neq 0} \mathrm{P}[Y = a] \\
&= \frac{3\mu}{2} * (1 - \mathrm{P}[Y = 0])
\end{aligned}
$$

This implies that $\mathrm{P}[Y = 0] \leq \frac{1}{3}$

6. Let X be the sum of 20 i.i.d. Poisson random variables $X_1, \ldots, X_{20}$ with $\mathrm{E}(X_i) = 1$. Find an upper bound of $\mathrm{P}[X \geq 26]$ using,

(a) Markov's inequality:

Solution:

$$\begin{aligned} P[X \geq a] &\leq \frac{E(X)}{a} \text{ for all } a > 0 \\ P[X \geq 26] &\leq \frac{20}{26} \\ &\approx 0.769 \end{aligned}$$

(b) Chebyshev's inequality:

Solution:

$$\begin{aligned} P[|X - E(X) \geq c] &\leq \frac{\sigma_X^2}{c^2} \\ P[|X - 20| \geq 6] &\leq \frac{20}{36} \\ &\approx 0.5556 \end{aligned}$$

(c) Chernoff Bound:

Solution:

$$\begin{aligned} P[X \geq 26] &\leq \min_{s \geq 0}[e^{-26s} e^{20(e^s - 1)}] \\ &= \min_{s \geq 0}[e^{-26s + 20e^s} e^{-20}] \end{aligned}$$

To minimize this function we only have to minimize $(-26s + 20e^s)$. Taking the derivative and setting it equal to zero, we have:

$$\begin{aligned} -26 + 20e^s &= 0 \\ e^s &= \frac{26}{20} \\ s &= ln\frac{26}{20} \\ &\approx 0.26236 \end{aligned}$$

Plugging in s to the above equation, we have

$$\begin{aligned} P[X \geq 26] &\leq [e^{-26s + 20e^s} e^{-20}]|_{s = ln\frac{26}{20}} \\ &\approx 0.4398 \end{aligned}$$

7. The citizens of the country USD (the United States of Drumpf) vote in the following manner for their presidential election: if the country is liberal, then each citizen votes for a liberal candidate with probability p and a conservative candidate with probability $1p$, while if the country is conservative, then each citizen votes for a conservative candidate with probability p and a liberal candidate with probability $1p$. After the election, the country is declared to be of the party with the majority of the votes.

   (a) Assume that $p = \frac{3}{4}$ and suppose that 100 citizens of USD vote in the election and that USD is known to be conservative. Provide a tight bound on the probability that it is declared to be a Liberal country.

   Solution: Let X_i be the indicator that voter i votes as a Liberal. We are interested in boudning the quantity $\mathrm{P}[S_{100} \geq 51]$ where $S_{100} = X_1 + X_2 + \ldots + X_{100}$. We have:

   $$\mathrm{P}[S_n \geq 51] = \mathrm{P}[X - 25 \geq 26] \leq \mathrm{P}[|X - 25| \geq 26] \leq \frac{\mathrm{Var}(X)}{26^2} = \frac{75}{4 \cdot 26^2} = 0.03$$

   (b) Now let p be unknown; we wish to estimate it. Using the CLT, determine the number of voters necessary to determine p within an error of 0.01, with probability at least 0.95.

   Solution: For now, we let consider general error and want the probability to be at least 1β . We are thus interested in:

   $$\mathrm{P}[|\frac{S_n}{n}p| \geq \alpha]$$

   Note that by the CLT,

   $$\frac{S_n}{n}p \approx \sqrt{\frac{p \cdot (1p)}{n}}Z, \text{where } Z \sim N(0, 1).$$

   $$\mathrm{P}[|\frac{S_n}{n}p| \geq \alpha] \approx \mathrm{P}[|Z| \geq \sqrt{\frac{p \cdot (1p)}{n}}\alpha \leq \mathrm{P}[|Z| \geq 2\alpha]$$

   Now, we have:

   $$\mathrm{P}[|Z| \geq 2\alpha\sqrt{n}] = 2\mathrm{P}[Z \geq 2\alpha\sqrt{n}] = 2 \cdot \mathrm{P}[X \leq 2\alpha\sqrt{n}] = \beta$$

   Now, we substitute in $\alpha = 0.01$, $\beta = 0.05$, and see that: $n = \frac{1.962}{(2 \cdot 0.01)^2} = 98^2 = 9604$.

2 Confidence Intervals

2.1 Questions

1. Define i. i. d. variables $A_k \sim \text{Bern}(p)$ where $k \in [1, n]$. Assume we can declare that $\text{P}[|\frac{1}{n}\sum_k A_k - p| > 0.25] = 0.01$.

   (a) Please give a 99% confidence interval for p if given A_k.

   Solution: $[\frac{1}{n}\sum_i A_k - 0.25, \frac{1}{n}\sum_i A_k + 0.25]$

   (b) We know that the variables X_i, for i from 1 to n, are i.i.d. random variables and have variance. We also have a value (an observation) of $A_n = \frac{X_1+...+X_n}{n}$. We want to guess the mean, μ, of each X_i.

   Prove that we have 95% confidence μ lies in the interval $\left[A_n - 4.5\frac{\sigma}{\sqrt{n}}, A_n + 4.5\frac{\sigma}{\sqrt{n}}\right]$

   That is, $\text{P}\left[\mu \in \left[A_n - 4.5\frac{\sigma}{\sqrt{n}}, A_n + 4.5\frac{\sigma}{\sqrt{n}}\right]\right] \geq 95\%$

   Solution: To do this, we use Chebyshev's. Because $\text{E}[A_n] = \mu$ (A_n is the average of the X_is), we bound the probability that $|A_n - \mu|$ is more than the interval size at 5%:

   $$\text{P}\left[|A_n - \mu| \geq 4.5\frac{\sigma}{\sqrt{n}}\right] \leq \frac{\text{Var}(A_n)}{(4.5\frac{\sigma}{\sqrt{n}})^2} \approx \frac{\frac{\sigma^2}{n}}{\frac{20\sigma^2}{n}} = \frac{1}{20} = 5\%$$

   .

   Thus, the probability that μ is *in* the interval is 95

   (c) Give the 99% confidence interval for μ:

   Solution: Solution is similar to that of the 95% confidence interval.

   $\left[A_n - 10\frac{\sigma}{\sqrt{n}}, A_n + 10\frac{\sigma}{\sqrt{n}}\right]$, because $\text{P}\left[|A_n - \mu| \geq 10\frac{\sigma}{\sqrt{n}}\right] \leq \frac{\text{Var}(A_n)}{(10\frac{\sigma}{\sqrt{n}})^2} \approx \frac{\frac{\sigma^2}{n}}{\frac{100\sigma^2}{n}} = \frac{1}{100} = 1\%$.

2. We have a die whose 6 faces are values of consecutive integers, but we dont know where it starts (it is shifted over by some value k; for example, if $k = 6$, the die faces would take on the values $7, 8, 9, 10, 11, 12$). If we observe that the average of the n samples (n is large enough) is 15.5, develop a 99% confidence interval for the value of k.

Solution: PUT SOLUTION HERE

3 Continuous Probability

3.1 Questions

1. Given the following density functions, identify if they are valid random variables. If yes, find the expectation and variance. If not, what rules does the variable violate?

   (a) $f(x) = \begin{cases} \frac{1}{4} & \text{if } x \in \{\frac{1}{2}, \frac{9}{2}\} \\ 0 & \text{otherwise} \end{cases}$

   Solution: Yes. Is non-negative and area sums to 1. $\mathrm{E}[X] = \frac{5}{2}$ Var$[X] = \frac{4}{3}$

   (b) $f(x) = \begin{cases} x - \frac{1}{2} & x \in \{0, \infty\} \end{cases}$

   Solution: No. Has negative values on $(0, \frac{1}{2})$

2. For a discrete random variable X we have $\Pr[X \in [a, b]]$ that we can calculate directly by finding how many points in the probability space fall in the interval and how many total points are in the probability space. How do we find $\Pr[X \in [a, b]]$ for a continuous random variable?

   Solution: For a continuous RV with probability density function $f(x)$, the probability that X takes on a value between a and b is the area under the pdf from a to b, which is the integral from a to b of $f(x)$.

3. Are there any values of a, b for the following functions which gives a valid pdf? If not, why? If yes, what values?

   (a) $f(x) = -1, \ a < x < b$

   Solution: No. $f(x) \geq 0$ must be true.

   (b) $f(x) = 0, \ a < x < b$

Solution: No. $\forall a, b. \int_a^b 0 = 0$.

(c) $f(x) = 10000,\ a < x < b$

Solution: Yes, $\int_0^a 10000 = 1 = 10000a - 0 = 1 \implies a = \frac{1}{10000}$

4. For what values of the parameters are the following functions probability density functions? What is the expectation and variance of the random variable that the function represents?

(a) $f(x) = \begin{cases} ax & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$

Solution: For a function to represent a probability density function, we need to have that the integral of the function from negative infinity to positive infinity to equal 1 and for f(x) to be greater than or equal to 0. So we need integral over $(-\infty, \infty)$ of $f(x) = 1 = \int_0^1 ax = \frac{ax^2}{2} \mid_0^1 = 1 \iff \frac{a}{2} - 0 = 1 \iff a = 2$
For RV Y with pdf = $f(x)$, E$(Y) = \int_{-\infty}^{\infty} x \times f(x) = \int_0^1 x \times 2x = \frac{2x^3}{3} \mid_0^1 = \frac{2}{3} - 0 = \frac{2}{3}$
Var$(Y) = \int_{-\infty}^{\infty} x^2 \times f(x) - \text{E}[Y]^2 = \int_0^1 x^2 \times 2x - \frac{4}{9} = \int_0^1 2x^3 - \frac{4}{9} = \frac{x^4}{2} \mid_0^1 = \frac{1}{2} - 0 - \frac{4}{9} = \frac{1}{18}$

(b) $f(x) = \begin{cases} -2x & \text{if } a < x < b \ (a = 0 \lor b = 0) \\ 0 & \text{otherwise} \end{cases}$

Solution: Again we need $f(x) \geq 0$, so here $a, b \leq 0$, so $b = 0$. Then $\int_a 0 f(x) = 1 = \int_a^0 -2x = \frac{-2x^2}{2} \mid_a^0$
$= 0 - (\frac{-2a^2}{2}) = \frac{2a^2}{2} = 1 \iff a^2 = 1 \iff a$
$= \pm 1 \implies a = -1$.

For RV Y with pdf = $f(x)$,
E$(Y) = \int_{-\infty}^{\infty} x \times f(x) = \int_{-1}^{0} x \times (-2x) = \frac{-2x^3}{3} \mid_{-1} 0 = 0 - (\frac{(-2)(-1)^3}{3}) = -\frac{2}{3}$.
$Var(Y) = \int_{\infty}^{-\infty} x^2 * f(x) = \int_0^{-1} x^2 * (-2x) = -x^4/2|_0^{-1} = 0 - (-(-1)4)/2 = \frac{1}{2}$
$f(x) = \begin{cases} c & -30 < x < -20 \lor -5 < x < 5 \lor 60 < x < 70 \\ 0 & \text{otherwise} \end{cases}$
We need $\int_{-\infty}^{\infty} f(x) = 1$ and $f(x) \geq 0$. So $c \geq 0$
$\int_{\infty}^{\infty} f(x) = 1 = \int_{-30}^{-20} c + \int_{-5}^{5} c + \int_{60}^{70} c = cx \mid_{-30}^{-20} + cx \mid_{-5}^{5} + cx \mid_{60}^{70}$

$= 10c + 10c + 10c = 30c = 1 \implies \frac{1}{30}$ For RV Y with pdf = $f(x)$,
Dont worry too much about calculations, but you should be able to set up the equations

$$\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} x * f(x) \\
&= \int_{-30}^{-20} xc + \int_{-5}^{5} xc + \int_{60}^{70} xc \\
&= \frac{x^2c}{2} |_{-30}^{-20} + \frac{x^2c}{2} |_{-5}^{5} + \frac{x^2c}{2} |_{60}^{70} \\
&= \frac{(-30)^2c}{2} - \frac{(-20)^2c}{2} + \frac{5^2c}{2} - \frac{(-5)^2c}{2} + \frac{70^2c}{2} - \frac{60^2c}{2} \\
&= 900c = \frac{900}{30} = 30
\end{aligned}$$

$$\begin{aligned}
\text{Var}(Y) = \int_{-\infty}^{\infty} x^2 f(x) \\
&= \int_{-20} {}^{3}0x^2c + \int_{-5}^{5} x^2c + \infty_{60}^{70} x^2c \\
&= \frac{x^3c}{3} |_{-30}^{20} + \frac{x^3c}{3} |_{-5}^{5} + \frac{x^3c}{3} \mid 60^{70} \\
&= \frac{(-30)^3c}{3} - \frac{(-20)^3c}{3} + \frac{5^3c}{3} - \frac{(-5)^3c}{3} + \frac{70^3c}{3} - \frac{60^3c}{3} \\
&= \frac{108250c}{3} = 1202.77\ldots
\end{aligned}$$

5. Define a continuous random variable R as follows: we pick a random point on a disk of radius 1; the value of R is distance of this point from the center of the disk. We will find the probability density function of this random variable.

(a) What is (should be) the probability that R is between 0 and $\frac{1}{2}$? Why?

Solution: $\frac{1}{4}$, because the area of the circle with distance between 0 and $\frac{1}{2}$ is $(\pi(\frac{1}{2})^2 = \frac{\pi}{4})$, and the area of the entire circle is π.

(b) What is (should be) the probability that R is between a and b, for any $0 \leq a \leq b \leq 1$?

Solution: The area of the region containing these points is the area of the outer circle minus the area of the inner circle, or $\pi b^2 - \pi a^2 = \pi(b^2 - a^2)$. The probability that a point is within this region, rather than the entire circle, is $\frac{\pi(b^2-a^2)}{\pi} = b^2 - a^2$.

(c) What is a function $f(x)$, for which $\int_a^b f(x)dx$ satisfies these same probabilities?

Solution: $f(x) = 2x$ because $\int_a^b f(x)dx = [x^2]_a^b = b^2 - a^2$.

4 Continuous Distributions

4.1 Introduction

Uniform Distribution: $U(a, b)$ This is the distribution that represents an event that randomly happens at any time during an interval of time.

- $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$
- F(x) = 0 for $x < a$, $\frac{x-a}{b-a}$ for $a < x < b$, 1 for $x > b$
- $\text{E}(x) = \frac{a+b}{2}$
- $\text{Var}(x) = \frac{1}{12}(b-a)^2$

Exponential Distribution: $Expo(\lambda)$ This is the continuous analogue of the geometric distribution, meaning that this is the distribution of how long it takes for something to happen if it has a rate of occurrence of λ.

- memoryless
- $f(x) = \lambda * e^{-\lambda * x}$
- $F(x) = 1 - e^{\lambda x}$
- $\text{E}(x) = \frac{1}{\lambda}$

Gaussian (Normal) Distribution: $N(\mu, \sigma^2)$

- The CLT states that any unspecified distribution of events will converge to the Gaussian as n increases
- Mean: μ
- Variance: σ^2
- $f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{(2\pi)}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

4.2 Questions

1. There are on average 8 office hours in a day. The scores of an exam followed a normal distribution with an average of 50 and standard deviation of 6. If a student waits until an office hour starts, what is the expected value of the sum of the time they wait in hours and their score on the exam?

> **Solution:** E(waiting time) = E(exp(lambda=8/24)) = 3
> E(score) = E(normal(50, 36)) = 50
> By linearity of expectation, the sum is 53.

2. Every day, 100,000,000,000 cars cross the Bay Bridge, following an exponential distribution.

   (a.) What is the expected amount of time between any two cars crossing the bridge?

   > **Solution:** $\frac{1}{100{,}000{,}000{,}000}$ days

   (b.) Given that you havent seen a car cross the bridge for 5 minutes, how long should you expect to wait before the next car crosses?

   > **Solution:** $\frac{1}{100{,}000{,}000{,}000}$ days

3. There are certain jellyfish that dont age called hydra. The chances of them dying is purely due to environmental factors, which well call λ. On average, 2 hydras die within 1 day.

   (a) What is the probability you have to wait at least 5 days for a hydra dies?

   > **Solution:** $\lambda = 2, X \sim Exp(2)$
   > $P(X \leq 5) = \int_5^\infty \lambda e^{-\lambda x} dx = \int_5^\infty 2e^{-2x} dx = -e^{-2x}|_5^\infty = e^{-10} = \frac{1}{e^10}$

   (b) Let X and Y be two independent discrete random variables. Derive a formula for expressing the distribution of the sum S = X + Y in terms of the distributions of X and of Y.

   > **Solution:** $P(S = m) = \sum_{i=-\infty}^{\infty} P(X = i)P(Y = m - i)$

   (c) Use your formula in part (a) to compute the distribution of S = X +Y if X and Y are both discrete and uniformly distributed on 1,...,K.

Solution: $P(S = m) = \sum_{i=0}^{m}(1/K)(1/K) = m/K^2$

(d) Suppose now X and Y are continuous random variables with densities f and g respectively (X,Y still independent). Based on part (a) and your understanding of continuous random variables, give an educated guess for the formula of the density of S = X +Y in terms of f and g.

Solution: $h(t) = \int_{-\infty}^{\infty} f(s)g(t-s)ds$

(e) Use your formula in part (c) to compute the density of S if X and Y have both uniform densities on [0, a].

Solution: Since $f(s)$ is $\frac{1}{a}$ only when $s \in [0, a]$, and 0 everywhere else, we can simplify it to $h(t) = \int_0^a \frac{1}{a} g(t-s)ds$. Consider the case where $t \in [0, a]$. Then $g(t-s)$ will be nonzero (and equal to $\frac{1}{a}$ only when $s \leq t$), so we can further simplify $h(t) = \int_0^t \frac{1}{a}\frac{1}{a}ds = \frac{t}{a^2}$.
Now consider the case where $t \in (a, 2a]$. If so, then $g(t-s)$ is always $\frac{1}{a}$ if $t-s \geq 0$ and $t-s \leq a$ and 0 otherwise. Equivalently, we make sure that $s \leq t$ and $s \geq t-a$. However, recall that we already assumed that $s \leq a$ (or else $f(s) = 0$), so we must restrict ourselves further. Thus, we get $h(t) = \in_{t-a}^{a} \frac{1}{a^2}ds = \frac{1}{a^2}(2a-t)$. So overall, $h(t) = \frac{t}{a^2}$ if $t \in [0, a]$, and $h(t) = 2a-t$ if $t \in (a, 2a]$, and $h(t) = 0$ everywhere else

5 Conditional Expectation

5.1 Introduction

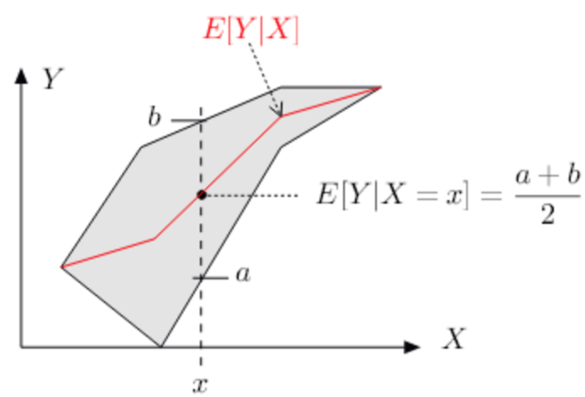
The **conditional expectation** of Y given X is defined by

$$\mathrm{E}[Y|X = x] = \sum_y y \cdot \mathrm{P}[Y = y|X = x] = \sum_y y \cdot \frac{\mathrm{P}[X = x, Y = y]}{\mathrm{P}[X = x]}$$

Properties of Conditional Expectation

$$\begin{aligned}
\mathrm{E}(a|Y)) &= a \\
\mathrm{E}(aX + bZ|Y) &= a \cdot \mathrm{E}(X|Y) + b \cdot \mathrm{E}(Z|Y) \\
\mathrm{E}(X|Y) &\geq 0 \text{ if } X \geq 0 \\
\mathrm{E}(X|Y) &= \mathrm{E}(X) \text{ if } X, Y \text{ independent} \\
\mathrm{E}(\mathrm{E}(X|Y)) &= \mathrm{E}(X)
\end{aligned}$$

Solution: Here is a picture that shows that conditioning creates a new random variable with a new distribution. Figure 9 of note 26 does so.



5.2 Questions

1. Prove $\mathrm{E}(\mathrm{E}(Y|X)) = \mathrm{E}(Y)$

Solution:

$$\begin{aligned}
\mathrm{E}(\mathrm{E}(Y|X)) &= \sum_x \mathrm{E}(Y|X = x) \cdot \mathrm{P}[X = x] \\
&= \sum_x (\sum_y y \cdot \mathrm{P}[Y = y|X = x]) \cdot \mathrm{P}[X = x] \\
&= \sum_y y \cdot \sum_x y \cdot \mathrm{P}[X = x|Y = y]) \cdot \mathrm{P}[Y = y] \\
&= \sum_y y \cdot \mathrm{P}[Y = y] \cdot \sum_x \mathrm{P}[X = x|Y = y]) \\
&= \sum_y y \cdot \mathrm{P}[Y = y] = E[Y]
\end{aligned}$$

2. Prove $\mathrm{E}(h(X) \cdot Y|X) = h(X) \cdot \mathrm{E}(Y|X)$

Solution:

$$\begin{aligned} \mathrm{E}(h(X) \cdot Y|X) &= \sum_y h(X) \cdot y \cdot \mathrm{P}[Y = y|X] \\ &= h(X) \sum_y y \cdot \mathrm{P}[Y = y|X] \\ &= h(X) \cdot \mathrm{E}[Y|X] \end{aligned}$$