Continuous Probability, Conditional Expectation 11

Computer Science Mentors 70

November 6 - November 10, 2017

1 Conditional Expectation

1.1 Introduction

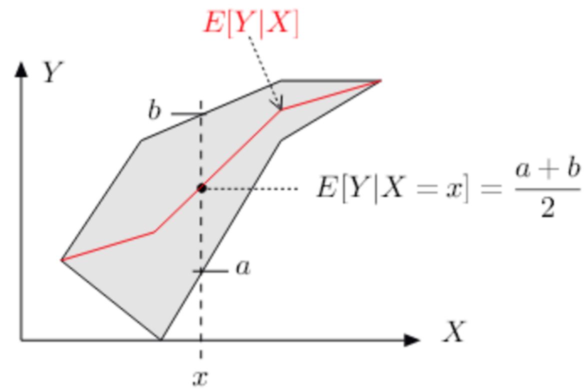
The **conditional expectation** of Y given X is defined by

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

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.



1.2 Questions

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

Solution:

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

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

Solution:

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

3. Consider the random variables Y and X with the following probabilities

   This table gives the probability distribution for $P[X \cap Y]$

		X		
		0	1	2
Y	0	0	.1	.2
	1	.1	.2	.1
	2	.2	.1	0

   Find:

   (a) $\mathrm{E}(Y|X=0)$

   Solution:

$$\begin{aligned}
\mathrm{E}(Y|X=0) &= \mathrm{P}[Y=0|X=0]\cdot 0 + \mathrm{P}[Y=1|X=0]\cdot 1 + \mathrm{P}[Y=2|X=0]\cdot 2 \\
&= \frac{0}{0+.1+.2}\cdot 0 + \frac{.1}{0+.1+.2}\cdot 1 + \frac{.2}{0+.1+.2}\cdot 2 \\
&= \frac{20}{12} = \frac{5}{3}
\end{aligned}$$

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

   Solution:

$$\begin{aligned}
\mathrm{E}(Y|X=1) &= \mathrm{P}[Y=0|X=1]\cdot 0 + \mathrm{P}[Y=1|X=1]\cdot 1 + \mathrm{P}[Y=2|X=1]\cdot 2 \\
&= \frac{0.1}{0.1+0.2+0.1}\cdot 0 + \frac{0.2}{0.1+0.2+0.1}\cdot 1 + \frac{0.1}{0.1+0.2+0.1}\cdot 2 \\
&= 0 + \frac{0.2}{0.4} + \frac{0.1\cdot 2}{0.4} \\
&= 0.5 + 0.5 = 1
\end{aligned}$$

   (c) $\mathrm{E}(Y|X=2)$

   Solution:

$$\begin{aligned}
\mathrm{E}(Y|X=2) &= \mathrm{P}[Y=0|X=2]\cdot 0 + \mathrm{P}[Y=1|X=2]\cdot 1 + \mathrm{P}[Y=2|X=2]\cdot 2 \\
&= \frac{0.2}{0.2+0.1+0}\cdot 0 + \frac{0.1}{0.2+0.1+0}\cdot 1 + \frac{0}{0.2+0.1+0}\cdot 2 \\
&= 0 + \frac{0.1}{0.3} + 0 = \frac{1}{3}
\end{aligned}$$

(d) $\mathrm{E}(Y)$

Solution: These events are disjoint, so to find $\mathrm{E}[Y]$, we can just sum up individual probabilities (note that sum of all probabilities is the sum of 1)

$$\begin{aligned}
\mathrm{E}(Y) &= \mathrm{E}(Y|X=0)\cdot \mathrm{P}[X=0] + \mathrm{E}(Y|X=1)\cdot \mathrm{P}[X=1] + \mathrm{E}(Y|X=2)\cdot \mathrm{P}[X=2] \\
&= \frac{20}{12}\cdot(0+0.1+0.2) + 1\cdot(0.1+0.2+0.1) + \frac{1}{3}\cdot(0.2+0.1+0) \\
&= \frac{20}{12}\cdot\frac{3}{10} + 0.4 + \frac{0.3}{3} \\
&= \frac{60}{120} + \frac{2}{5} + \frac{1}{10} \\
&= \frac{60}{120} + \frac{48}{120} + \frac{12}{120} = \frac{120}{120} = 1
\end{aligned}$$

2 Continuous Probability

2.1 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. $\mathrm{E}[X] = \frac{5}{2}$ $\mathrm{Var}[X] = \frac{4}{3}$

   (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) $= \int_{-\infty}^{\infty} x^2 \times f(x) - \mathrm{E}[Y]^2 = \int_0^1 x^2 \times 2x - \frac{4}{9} = \int_0^1 2x^3 - \frac{4}{9} = \frac{x^4}{2} \mid_0^1 = \frac{1}{2} - 0 - \frac{4}{9} = \frac{1}{18}$

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

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 = $f(x)$,
$\text{E}(Y) = \int_{-\infty}^{\infty} x \times f(x) = \int_{-1}^{0} x \times (-2x) = \frac{-2x^3}{3} \mid_{-1} 0 = 0 - (\frac{(-2)(-1)^3}{3}) = -\frac{2}{3}.$
$Var(Y) = \int_{\infty}^{-\infty} x^2 * f(x) = \int_0^{-1} x^2 * (-2x) = -x^4/2|_0^{-1} = 0 - (-(-1)4)/2 = \frac{1}{2}$
$f(x) = \begin{cases} c & -30 < x < -20 \vee -5 < x < 5 \vee 60 < x < 70 \\ 0 & \text{otherwise} \end{cases}$
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

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

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

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

3 Continuous Distributions

3.1 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 $x > b$
- $\mathrm{E}(x) = \frac{a+b}{2}$
- $\mathrm{Var}(x) = \frac{1}{12}(b-a)^2$

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

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

3.2 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: Since the student and the office hours lie on the same distribution, there is no difference between when the student comes and when the office hours are held. So

E(waiting time) = E(exp(lambda=9/24)) = 8/3
E(score) = E(normal(50, 36)) = 50
By linearity of expectation, the sum is $52\frac{2}{3}$.

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

   (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