CSM 70 Discrete Mathematics and Probability Theory

Spring 2016 Intro to Continuous, Markov and Chebyshev **Worksheet 13 Solutions**

I. Integral/Expectation/Analogous things

Given some different density functions, are these valid RVs? If yes, find expectation and Variance. If not, what rules does it violate?
Given some expectation and the bounds, find the bounds of the variable

1. Given the following density functions, identify if they are valid random variables. If yes, find the expectation and variance. If no, what rules does the variable violate?
   a. f(x) = ¼ on (½, 9/2), = 0 elsewhere
      i. Expectation: 5/2
      ii. Variance: 4/3
   b. f(x) = x - ½ on (0, $$\infty$$)
      i. Has negative values on (0, ½)
2. For a discrete random variable X we have Pr(X within [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 within [a, b]) for a continuous random variable?

   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 which we have a valid pdf? If not, why? If yes, what values?

   f(x) = -1 a<x<b

   No. f(x) >= 0 must be true.

   f(x) = 0 a<x<b (Are there any values of a,b for which we have a valid pdf?)

   No. S^b_a 0 = 0 for all a,b.

   f(x) = 10000, 0<x<a (Are there any values of a for which we have a valid pdf?)

   Yes, S^a_0 10000 = 1 = 10000a - 0 = 1 ⇒ a = 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?

   f(x) = ax, 0<x<1, f(x) = 0 otherwise

   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 0 and for f(x) to be greater than or equal to 0. So we need integral over (-inf, inf) of f(x) = 1
   = integral over (0, 1) of ax
   = $(ax^2)/2 |_0^1 = 1$ ⇔ a/2 - 0 = 1 ⇔ a = 2

   For RV Y with pdf = f(x),

   E(Y) = integral over (-inf, inf) of x*f(x)
   =integral over(0,1) of x*(2x) = $2x^3/3 |^1_0$ = ⅔ - 0 = ⅔

   Var(Y) = integral over (-inf, inf) of x^2*f(x) - E(Y) = integral over (0, 1) of x^2*2x
   = integral over(0,1) of $2x^3 = x^4/2 |^1_0$ = ½ - 0 = ½

$f(x) = -2x$, $a<x<b$, ($a=0$ OR $b=0$), $f(x) = 0$ otherwise

Again we need $f(x) >= 0$, so here $a, b <= 0$, so $b=0$.

Then $S^0_a\ f(x) = 1 = S^0_a\ -2x = -2x^2/2\ |^0_a = 0 - (-2a^2/2) = 2a^2/2 = 1$

$\Leftrightarrow a^2 = 1 \Leftrightarrow a = +- 1 \Rightarrow a = -1$

For RV Y with pdf = f(x),

$E(Y) = S^{inf}_{-inf}\ x*f(x) = S^0_{-1}\ x*(-2x) = -2x^3/3\ |^0_{-1} = 0 - (-2(-1)^3/3) = -⅔$

$Var(Y) = S^{inf}_{-inf}\ x^2*f(x) = S^0_{-1}\ x^2*(-2x) = -x^4/2\ |^0_{-1} = 0 - (-(-1)^4)/2 = ½$

$f(x) = c$, $-30<x<-20$, $-5<x<5$, $60<x<70$, $f(x) = 0$ otherwise

We need $S^{inf}_{-inf}\ f(x) = 1$ and $f(x) >= 0$. So $c>=0$.

$S^{inf}_{-inf}\ f(x) = 1 = S^{-20}_{-30}\ c + S^5_{-5}\ c + S^{70}_{60}\ c = cx|^{-20}_{-30} + cx|^5_{-5} + cx|^{70}_{60}$

$= 10c + 10c + 10c = 30c = 1 \Rightarrow c = 1/30$

For RV Y with pdf = f(x),

Don't worry too much about calculations, but you should be able to set up the equations

$E(Y) = S^{inf}_{-inf}\ x*f(x) = S^{-20}_{-30}\ xc + S^5_{-5}\ xc + S^{70}_{60}\ xc = x^2c/2|^{-20}_{-30} + x^2c/2|^5_{-5} + x^2c/2|^{70}_{60}$

$= (-30)^2c/2 - (-20)^2c/2 + 5^2c/2 - (-5)^2c/2 + 70^2c/2 - 60^2c/2 = 900c = 900/30 = 30$

$Var(Y) = S^{inf}_{-inf}\ x^2f(x) = S^{inf}_{-inf}\ x^2*f(x) = S^{-20}_{-30}\ x^2c + S^5_{-5}\ x^2c + S^{70}_{60}\ x^2c$

$= x^3c/3|^{-20}_{-30} + x^3c/3|^5_{-5} + x^3c/3|^{70}_{60} = (-30)^3c/3 - (-20)^3c/3 + 5^3c/3 - (-5)^3c/3 + 70^3c/3 - 60^3c/3$

$= 108250c/3 = 1202.77…$

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 ½? Why?
      ¼, because the area of the circle with distance between 0 and ½ is ($\pi(½)^2 = \pi/4$), and the area of the entire circle is
   b. What is (should be) the probability that R is between *a* and *b*, for any $0 \leq a \leq b \leq 1$?
      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 $\pi(b^2 - a^2) / \pi = b^2 - a^2$.
   c. What is a function f(*x*), for which $\int_a^b f(x)dx$ satisfies these same probabilities?
      $f(x) = 2x$, because $\int_a^b f(x)dx = \left[x^2\right]_a^b = b^2 - a^2$
   d. Define *g*(*x*), the probability density function for R.
      $$g(x) = \begin{cases} 2x, & \text{if } x \in [0,1] \\ 0, & \text{otherwise} \end{cases}$$

II. Distributions (Uniform, Exponential, Gaussian=normal, Zipf) (longest)

- Mechanical problem for Gaussian + Exponential
- Problem exploiting memorylessness of Exp
- Prove memorylessness of Exp

There are certain jellyfish that don't age called hydra. The chances of them dying is purely due to environmental factors, which we'll call *lambda*. 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?

$\lambda = 2$, $X \sim Exp(2)$, $P(X \geq 5) = \int\limits_5^\infty \lambda e^{-\lambda x}dx = \int\limits_5^\infty 2e^{-2x}dx = -e^{-2x}|_5^\infty = e^{-10} = \frac{1}{e^{10}}$

(a) 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.

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

(b) 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 $\{1,...,K\}$.

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

(c) 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.

$$h(t) = \int\limits_{-\infty}^{\infty} f(s)g(t - s)ds$$

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

$$h(t) = \int\limits_{-\infty}^{\infty} f(s)g(t - s)ds$$

Since $f(s)$ is $\frac{1}{a}$ only when $s \in [0, a]$, and 0 everywhere else, we can simplify it to $h(t) = \int\limits_0^a \frac{1}{a}g(t - s)ds$

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 simply $h(t) = \int\limits_0^t \frac{1}{a}\frac{1}{a}ds = \frac{t}{a^2}$

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 $h(t) = \int\limits_{t-a}^{a} \frac{1}{a^2}ds = \frac{1}{a^2}(2a - t)$.

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.

(e) Show that if *X* and *Y* are independent normally distributed variables, then *X* + *Y* is also a normally distributed variable.

For independent random variables X and Y, the distribution f_Z of $Z = X+Y$ equals the convolution of f_X and f_Y:

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(z-x) f_X(x) dx$$

Given that f_X and f_Y are normal densities,

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

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}$$

Substituting into the convolution:

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{(z-x-\mu_Y)^2}{2\sigma_Y^2}} \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2+\sigma_Y^2}} \exp\left[-\frac{(z-(\mu_X+\mu_Y))^2}{2(\sigma_X^2+\sigma_Y^2)}\right] \frac{1}{\sqrt{2\pi}\frac{\sigma_X\sigma_Y}{\sqrt{\sigma_X^2+\sigma_Y^2}}} \exp\left[-\frac{\left(x-\frac{\sigma_X^2(z-\mu_Y)+\sigma_Y^2\mu_X}{\sigma_X^2+\sigma_Y^2}\right)^2}{2\left(\frac{\sigma_X\sigma_Y}{\sqrt{\sigma_X^2+\sigma_Y^2}}\right)^2}\right] dx$$

$$= \frac{1}{\sqrt{2\pi(\sigma_X^2+\sigma_Y^2)}} \exp\left[-\frac{(z-(\mu_X+\mu_Y))^2}{2(\sigma_X^2+\sigma_Y^2)}\right] \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\frac{\sigma_X\sigma_Y}{\sqrt{\sigma_X^2+\sigma_Y^2}}} \exp\left[-\frac{\left(x-\frac{\sigma_X^2(z-\mu_Y)+\sigma_Y^2\mu_X}{\sigma_X^2+\sigma_Y^2}\right)^2}{2\left(\frac{\sigma_X\sigma_Y}{\sqrt{\sigma_X^2+\sigma_Y^2}}\right)^2}\right] dx$$

The expression in the integral is a normal density distribution on x, and so the integral evaluates to 1. The desired result follows:

$$f_Z(z) = \frac{1}{\sqrt{2\pi(\sigma_X^2+\sigma_Y^2)}} \exp\left[-\frac{(z-(\mu_X+\mu_Y))^2}{2(\sigma_X^2+\sigma_Y^2)}\right]$$

III. Joint Distributions (not this week)

1. Suppose X and Y are independent and X has probability density function g(x) = 6x(1-x) for 0 <= x <=1, and that Y has probability density function h(y) = 12y^2(1-y) for 0 <= y <=1.

a. Find the probability density function of (X,Y)

b. Find P(X+Y <= 1)

Solutions:

   a. 72x(1-x)y^2(1-y)
   b. 13/35

2. Suppose that (X,Y) has probability f(x,y) = 6x^2y for x and y in [0,1].
   a. Find the probability of density function of X.
   b. Find the pdf.
   c. Are X and Y independent.

Solutions:

   a. 3x^2
   b. 2y
   c. Yes

TODO: WORD PROBLEM FOR NEXT WEEK

10. Choose m real numbers at random between [0,1]. Let X be the largest one of these numbers. Find the pdf of f(x) and the E[X].

ANSWER: We know that the Pr(X<= x) = x^m. From here, we can take the derivative and get mx^{m-1} as our pdf, or f(x). Now, to find the E[X], we can take the integral from 0 to 1 of x * f(x), and we get m / (m+1) as our expectation.

11. Intuitively, find the value that minimizes (E[(X-a)^2]), then prove your answer.

Answer: a should E[X] in order to get the deviation to be around 0. Now, mathematically, we can expand the inner term and get:
E[X^2] - 2aE[X] + a^2, which, when trying to minimize we will take the derivative and get -2E[X] + 2a = 0, so a must be equal to E[X]..