Chapter 3 Summary and Review (draft: 2019/10/02-02:25:54)

Random Variables

Summary of topics and terminology:

- Given sample space S, a (real-valued) random variable is a function $X : S \to \mathbb{R}$.
- Be able to calculate probabilities for both discrete and continuous random variables.
- A random variable is called discrete if it takes on a finite or countable number of values. E.g. $X \in \{1, 2, 3, 4, 5, 6\}$ or $X \in \mathbb{Z}$.
- A random variable is called continuous if it takes values in an interval, e.g. $X \in [0, 1]$ or $X \in (0, \infty)$.

Example problems:

1. Flip a fair coin 3 times, let X be the number of heads minus the number of tails. Find $P(|X| = 1)$.

   Solution: $X = \pm 1$ means that we have 2 H's and 1 T or vice versa. There are 3 ways to get each: 2H-1T: HHT, HTH, THH, or 1H-2T: HTT, THT, TTH. There are a total of $2^3 = 8$ equally likely outcomes, thus $P(|X| = 1) = 6/8$.

2. Consider a bag with 4 red balls and 3 blue balls. Draw 5 w/o replacement. Let X be the number of blue balls drawn. Find all possible values for X and the probabilities of each (the pmf).

   Solution:

   $P(X = x) = \dfrac{\binom{4}{5-x}\binom{3}{x}}{\binom{7}{5}}$ for $x = 1, 2, 3$.

   Carefully think about the allowable values of X. there must be at least 1 blue ball drawn wince we are drawing 5 and there are only 4 red balls. Of course, we cannot draw more than 3 blue balls.

   Let's check the law of total probability: we want $P(X = 1) + P(X = 2) + P(X = 3) = 1$.

   $$\frac{\binom{4}{4}\binom{3}{1}}{\binom{7}{5}} + \frac{\binom{4}{3}\binom{3}{2}}{\binom{7}{5}} + \frac{\binom{4}{2}\binom{3}{3}}{\binom{7}{5}} = \frac{1}{7} + \frac{4}{7} + \frac{2}{7} = 1$$

3. Consider the experiment where we flip a fair coin until we get a head. Let X be teh total number of flips. Is this a random variable? What are the probabilities for its values?

   Solution:

   $P(X = k) = 1/2^k$ for $k = 1, 2, 3, \ldots$.

   There is at least 1 flip ($k = 1$), but we can't limit it to a maximum finite number of flips. But the probability decays exponentially ast eh number of flips gets very large. We are unlikely to get a large number of tails without any heads!

   Let's check the law of total probability: $\sum_{k=1}^{\infty} 1/2^k = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = \frac{1}{2} \cdot \frac{1}{1-\frac{1}{2}} = 1$. This is a geometric series. (*typo fixed)

Univariate probability functions

Summary of topics and terminology:

- pmf = probability mass function (gives probability directly), $f(x) = P(X = x)$
- pdf = probability density function (integrate to get probability), $f(x) = \frac{d}{dx}P(X \leq x)$.
- cdf = cumulative distribution function $F(x) = P(X \leq x)$.
- $f(x) = F'(x)$ for continuous RV
- for discrete RV, cdf is a step function, height of step gives probability mass for value where step up occurs
- $P(X = x) = 0$ for continuous RV
- $P(a < X < b) = \int_a^b f(x)dx$ for a continous RV. The inequalities can be any combination of strict or not. This is not true for a discrete random variable.
- $P(a < X \leq b) = F(b) - F(a)$ for a discrete random variable. The inequalities must be exactly as is here otherwise care must be taken.
- "Uniformly distributed" means the pmf or pdf is a constant function. E.g. $f(x) = 1/(b-a)$ is $X \sim U[a, b]$ (continuous), and $f(x) = 1/n$ for $X \sim U\{x_1, x_2, ..., x_n\}$ (discrete).

Example problems:

1. Consider pmf $f(x) = x/6$ for $x = 1, 2, 3$. Write the formula for the cdf.

   Solution:

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1/6 & \text{if } 1 \leq x < 2 \\ 3/6 & \text{if } 2 \leq x < 3 \\ 1 & \text{if } x \geq 3 \end{cases}$$

2. Consider cdf given below. Find $P(2 \leq X < 5)$.

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ 0.05 & \text{if } 1 \leq x < 2 \\ 0.12 & \text{if } 2 \leq x < 3 \\ 0.3 & \text{if } 3 \leq x < 4 \\ 0.4 & \text{if } 4 \leq x < 5 \\ 0.73 & \text{if } 5 \leq x < 6 \\ 0.92 & \text{if } 6 \leq x < 7 \\ 1 & \text{if } x \geq 7 \end{cases}$$

   Solution:

   $P(2 \leq X < 5) = P(1 < X \leq 4) = F(4) - F(1) = 0.4 - 0.05 = 0.35$ **(* typo fixed here)**

   For discrete random variables, you need to pay careful attention to which values are actually included.

   edit: Here are a few variations on the inequalities:
   $P(2 \leq X \leq 5) = P(1 < X \leq 5) = F(5) - F(1) = 0.73 - 0.05 = 0.68$
   $P(2 < X < 5) = P(2 < X \leq 4) = F(4) - F(2) = 0.4 - 0.12 = 0.28$
   $P(2 < X \leq 5) = F(5) - F(2) = 0.73 - 0.12 = 0.61$

3. Given cdf $F(x) = 1 - \frac{1}{x^2}$ for $x > 0$, find the pdf.

   Solution:

   $f(x) = F'(x) = \frac{2}{x^3}$. **(*typo fixed)**

4. Given pdf $f(x) = 1/2$ on $[1, 2)$ and $f(x) = 3 - x$ on $[2, 3]$, find the cdf.

   Solution:

   $F(x) = 0$ if $x < 1$.

   $F(x) = P(X \leq x) = \int_1^x 1/2 \; dx = (x-1)/2$ if $x \in [1, 2)$

   $F(x) = 1/2 + \int_2^x (3-x) \; dx = 3(x-2) - \frac{1}{2}(x^2 - 4) = -\frac{1}{2}(x-3)^2 + 1$ if $x \in [2, 3]$

   $F(x) = 1$ for $x > 3$.

Multivariate probability functions

Summary of topics and terminology:

- Joint cdf 2 variables: $F(x,y) = P(X \leq x, Y \leq y)$. Note that this is an *and* statement, or the probability of the intersection of events $\{X \leq x\}$ and $\{Y \leq y\}$.
- Joint cdf for n variables: $F(x_1, x_2, ..., x_n) = P\left(\bigcap_{i=1}^{n} \{X_i \leq x_i\}\right)$.
- Joint pdf 2 variables: $f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$
- Joint pdf n variables: $f(x_1, x_2, ..., x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F(x,y)$
- $F(x,y) = \displaystyle\int_{-\infty}^{y} \int_{-\infty}^{x} f(t,s)\ dt\ ds$, but be careful that $f(x,y)$ might be zero in some places. More generally, this works for any number of variables, we just have as many integrals as there are variables.
- Joint cdf for discrete $F(x,y) = \displaystyle\sum_{s \leq y} \sum_{t \leq x} f(t,s)$, again be careful that $f(x,y)$ might be zero in some places. We can think of it like this : $F(x,y) = \sum_{\{(t,s): t \leq x, s \leq y\}} f(t,s)$.
- Marginal pdf for X: $f_X(x) = \displaystyle\int_{\mathbb{R}} f_{X,Y}(x,y)\ dy$. In book stated as $g(x) = \displaystyle\int_{-\infty}^{\infty} f(x,y)\ dy$.
- Marginal pdf for Y: $f_Y(y) = \displaystyle\int_{\mathbb{R}} f_{X,Y}(x,y)\ dx$. In book stated as $h(y) = \displaystyle\int_{-\infty}^{\infty} f(x,y)\ dx$.
- Conditional pdf of X given a fixed Y value: $f_{X|Y=y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$. In our textbook it is stated as $f(x|y) = \frac{f(x,y)}{h(y)}$. It is required that the denominator be nonzero.
- Conditional pdf for Y given a fixed X value: $f_{Y|X=x}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$. In the textbook this is stated as $w(y|x) = \frac{f(x,y)}{g(x)}$. It is required that the denominator be nonzero.
- For joint pdf $f(x,y,z)$ the marginal pdf for X is given by $f_X(x) = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y,z)\ dy\ dz$. We are integrating out both Y and Z. Other individual marginal pdfs can be calculated similarly.
- For joint pdf $f(x,y,z)$ the marginal joint pdf for X and Y is given by $f_{X,Y}(x,y) = \int_{\mathbb{R}} f(x,y,z)\ dz$. We are just integrating out Z. Other individual marginal pdfs can be calculated similarly.
- For joint pdf $f(x,y,z)$ the joint conditional pdf for X and Y given $Z = z$ is given by $f_{X,Y|Z=z}(x,y|z) = \frac{f(x,y,z)}{f_Z(z)}$.
- For joint pdf $f(x,y,z)$ the joint conditional pdf for X given $Y = y$ and $Z = z$ is given by $f_{X|Y=y,Z=z}(x|y,z) = \frac{f(x,y,z)}{f_{Y,Z}(y,z)}$.

Example problems:

1. Consider the experiment where 3 balls are randomly drawn without replacement out of an urn containing 4 red, 5 green, and 2 blue balls. Let X be the number of red balls and Y be the number of blue balls that are included in those drawn out. Write the joint pmf for X and Y. Find the marginal pmf of X. And find the conditional pmf for Y given that $X = 1$.

   Solution:

   Joint pmf:

$$f_{X,Y}(x,y) = \frac{\binom{4}{x}\binom{5}{3-x-y}\binom{2}{y}}{\binom{11}{3}}$$

   There is a domain restriction though since we require $1 \leq x + y \leq 3$

   Marginal pmf for X: $f_X(x) = f_{X,Y}(x,0) + f_{X,Y}(x,1) + f_{X,Y}(x,2) = \frac{\binom{4}{x}\binom{5}{3-x}\binom{2}{0}}{\binom{11}{3}} + \frac{\binom{4}{x}\binom{5}{2-x}\binom{2}{1}}{\binom{11}{3}} + \frac{\binom{4}{x}\binom{5}{1-x}\binom{2}{2}}{\binom{11}{3}}$ for $x = 0, 1, 2, 3$

   Conditional pmf

$$\begin{aligned}
f_{Y|X=1}(y|1) &= \frac{f_{X,Y}(1,y)}{f_X(1)} \\
&= \frac{\frac{\binom{4}{1}\binom{5}{3-1-y}\binom{2}{y}}{\binom{11}{3}}}{\frac{\binom{4}{1}\binom{5}{3-1}\binom{2}{0}}{\binom{11}{3}} + \frac{\binom{4}{1}\binom{5}{2-1}\binom{2}{1}}{\binom{11}{3}} + \frac{\binom{4}{x}\binom{5}{1-1}\binom{2}{2}}{\binom{11}{3}}} \\
&= \frac{4\binom{5}{2-y}\binom{2}{y}}{4\binom{5}{2}\binom{2}{0} + 4\binom{5}{1}\binom{2}{1} + 4\binom{5}{0}\binom{2}{2}} \\
&= \frac{\binom{5}{2-y}\binom{2}{y}}{\binom{5}{2}\binom{2}{0} + \binom{5}{1}\binom{2}{1} + \binom{5}{0}\binom{2}{2}} \\
&= \frac{\binom{5}{2-y}\binom{2}{y}}{10 + 10 + 1}
\end{aligned}$$

   for $y = 0, 1, 2$. We can check this by plugging in these y-values and adding the resulting conditional probabilities to make sure they sum to 1.

   This problem may be more easily understood writing out eh join pmf in tabular format:

		X				
		0	1	2	3	total
	0	10/165	40/165	30/165	4/165	84/165
Y	1	20/165	40/165	12/165	0	72/165
	2	5/165	4/165	0	0	9/165
	total	35/165	84/165	42/165	4/165	1

   If we look at the column $X = 1$ then the probabilities for each Y value are $40/165, 40/165, 4/165$ but we need to normalize these by dividing by their sum and this gives probabilities $40/84, 40/84, 4/84$. These simplify to $10/21, 10/21, 1/21$ for $y = 0, 1, 2$. This is the conditional pmf for Y given that $X = 1$.

   The marginal pmf for X are the column totals $35/165, 84/165, 42/165, 4/165$ for $X = 0, 1, 2, 3$.

2. Find the constant k and find the joint cdf for joint pdf given by $f(x,y) = \frac{1}{k}(3x^2+2y)$ on $[0,1]\times[0,2]$.

<u>*Solution:*</u>

We require $\int f(x,y)dA = 1$ and calculate $\int_0^1\int_0^2 \frac{1}{k}(3x^2+2y)dydx = \int_0^1 \left[\frac{1}{k}(3x^2y+y^2)\right]_0^2 dx = \int_0^1 \frac{1}{k}(6x^2+4)dx = \frac{1}{k}(2x^3+4x)\Big|_0^1 = \frac{6}{k}$ so $k=6$.

To find the cdf, we integrate over several different regions carefully. If x or y are below their allowable ranges, the cdf is zero: $F(x,y)=0$ if $x<0$ or $y<0$. When both x and y are within their allowable ranges we integrate: $F(x,y) = \int_0^y\int_0^x \frac{1}{6}(3t^2+2s)dtds = \frac{1}{6}(x^3y+xy^2)$ when $x\in[0,1]$ and $y\in[0,2]$. Now when $y>2$, we must stop the y-integral at $y=2$ since the pdf is zero beyond that. Similarly, we stop integrating if the x value goes beyond 1 since the pdf is zero beyond that. For $x\in[0,1]$ and $y>2$ we have $F(x,y) = \int_0^2\int_0^x \frac{1}{6}(3t^2+2s)dtds = \frac{1}{6}(2x^3+4x)$. For $x>1$ and $y\in[0,2]$ we have that $F(x,y) = \int_0^y\int_0^1 \frac{1}{6}(3t^2+2s)dtds = \frac{1}{6}(y+y^2)$. We summarize this as

$$F(x,y) = \begin{cases} 0 & \text{if } x<0 \text{ or } y<0 \\ \frac{1}{6}(x^3y+xy^2) & \text{if } 0\le x\le 1 \text{ and } 0\le y\le 2 \\ \frac{1}{6}(2x^3+4x) & \text{if } 0\le x\le 1 \text{ and } y>2 \\ \frac{1}{6}(y+y^2) & \text{if } x>1 \text{ and } 0\le y\le 2 \end{cases}$$

See https://www.math3d.org/PFgHn18m for a 3D plot of this cdf.

3. Continuing the previous problem, find

   (a) $P(X<0.5, Y>1)$

   (b) $P(X+Y<1)$

<u>*Solution:*</u>

(a) $P(X<0.5,Y>1) = \int_0^{0.5}\int_1^2 f(x,y)\,dy\,dx = \frac{1}{6}\left(x^3\big|_0^{0.5}(2-1) + y^2\big|_1^2\right) = \frac{17}{48}$

(b) **(*typo fixed)**

$$\begin{aligned} P(X+Y<1) &= P(X<1-Y) \\ &= \frac{1}{6}\int_0^1\int_0^{1-y} f(x,y)\,dx\,dy \\ &= \frac{1}{6}\int_0^1 \left[x^3+2yx\right]_0^{1-y}\,dy \\ &= \frac{1}{6}\int_0^1 (1-y)^3+2y(1-y)\,dy \\ &= \frac{1}{6}\left(-\frac{1}{4}(1-y)^4+y^2-\frac{2}{3}y^3\Big|_0^1\right) \\ &= \frac{1}{6}\left(\frac{1}{4}+1-\frac{2}{3}\right) = \frac{7}{72} \end{aligned}$$

Note that the y-limits stop at 1 since $X<1-Y$ is no longer possible if $Y>1$. This is easier to see by graping the region.

4. Continuing the previous problem again, find the marginal pdfs for X and for Y, and find the conditional pdfs for X given $Y = 1$ and for Y given $X = 0.5$.

   Solution:

   We have joint pdf $f(x, y) = \frac{1}{6}(3x^2 + 2y)$ on $[0, 1] \times [0, 2]$. We integrate out each variable individually to get the marginal distributions first:

   $f_X(x) = \int_0^2 f(x, y)dy = \frac{1}{6}(3x^2y + y^2)\big|_0^2 = \frac{1}{6}(6x^2 + 4)$ for $x \in [0, 1]$

   $f_Y(y) = \int_0^1 f(x, y)dx = \frac{1}{6}(x^3 + 2xy)\big|_0^1 = \frac{1}{6}(1 + 2y)$ for $y \in [0, 2]$.

   It is important to realize that these marginal pdfs are defined as zero outside of the domains given.

   The conditional pdfs are:

   $f_{X|Y=y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{3x^2+2y}{1+2y}$ for any *fixed* y-value in $[0, 2]$.
   Thus $f_{X|Y=1}(x|1) = \frac{1}{3}(3x^2 + 2)$.

   $f_{Y|X=x}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{3x^2+2y}{6x^2+4}$ for any *fixed* x-value in $[0, 1]$.
   Thus $f_{Y|X=0.5}(y|0.5) = \frac{2}{11}\left(\frac{3}{4} + 2y\right)$.

   Note: you should check that the integrals of these conditional pdfs over their respective domains satisfies the law of total probability.

5. Consider joint pdf $f_{X,Y,Z}(x, y, z) = \frac{1}{14}(x^2y + y^2z)$ on $[0, 1] \times [0, 2] \times [0, 3]$. Find the marginal pdfs $f_{X,Y}(x, y)$ and $f_Z(z)$. Find the conditional pdfs $f_{X,Y|Z=z}(x, y|z)$ and $f_{Z|X=x,Y=y}(z|x, y)$.

   Solution:

   $$f_{X,Y}(x, y) = \int_0^3 \frac{1}{14}(x^2y + y^2z)\ dz = \frac{1}{14}\left(3x^2y + \frac{9}{2}y^2\right)$$

   $$f_Z(z) = \int_0^2 \int_0^1 \frac{1}{14}(x^2y + y^2z)\ dx\ dy = \frac{1}{21}(1 + 4z)$$

   $f_{X,Y|Z=z}(x, y|z) = \frac{f_{X,Y,Z}(x,y,z)}{f_Z(z)} = \frac{\frac{1}{14}(x^2y+y^2z)}{\frac{1}{21}(1+4z)} = \frac{3}{2}\frac{x^2y+y^2z}{1+4z}$ (for a fixed value of z) **(*typo fixed)** here is a3D plot of the pdf with the plane cutting part of it off:
   https://www.math3d.org/yWyxL6HP
   You can visualize that the small triangular wedge that is cut off is fairly close to 1/10 the total volume.

   $f_{Z|X=x,Y=y}(z|x, y) = \frac{f_{X,Y,Z}(x,y,z)}{f_{X,Y}(x,y)} = \frac{\frac{1}{14}(x^2y+y^2z)}{\frac{1}{14}\left(3x^2y+\frac{9}{2}y^2\right)} = \frac{x^2y+y^2z}{3x^2y+\frac{9}{2}y^2}$. (For fixed values of x and y.)

   For example, if we wanted to find the conditional distribution of Z given $X = \frac{1}{\sqrt{3}}$ and $Y = 2$, then $f_{Z|X=\sqrt{3},Y=2}(z|\sqrt{3}, 2) = \frac{1}{10}(\frac{1}{3} + 2z)$.