Chapter 4

- We are now moving to multivariate distributions from univariate distributions more is better, right?

Scene 1: 1 datatype multiple observations
Scene 2: many datatypes multiple observations

- Don't worry we'll take it slow by learning bivariate at first

DISCRETE

Definition 4.1.1 - An n-dimensional random vector is a function from a sample space S into $\mathbb{R}^n$, an n-dimensional Euclidean Space

ie $\forall$ observation we have an ordered pair $(x,y) \in \mathbb{R}^2$ where $\mathbb{R}^2$ denotes a plane
This is an example of a bivariate random vector.

Example: The random experiment of rolling two fair dice
$\exists$ $6 \times 6 = 36$ possible outcomes
$\mathcal{B}_1 = \{(1,1), (1,2) \ldots (6,6)\}$ • note each outcome $\in \mathbb{R}^2$
• Assume that these are ordered pairs
ie: $(4,1) \neq (1,4)$

• We can consider X = sum of two dice Y = |difference of dice|
• We can then consider the bivariate RV (X,Y)
say for (4,1) $X = 4+1 = 5$ $Y = |4-1| = 3 \rightarrow (5,3)$
$\mathcal{B}_2 = \{(2,0), (2,1), \ldots (12,5)\}$ • note each outcome $\in \mathbb{R}^2$
• These are ordered (x,y)

$$P(X=x \text{ and } Y=y) = \frac{\text{outcomes in } \mathcal{B}_1 \text{ that satisfies these properties}}{36}$$

$$P(X=5 \text{ and } Y=3) = \frac{(4,1),(1,4)}{36} = \frac{2}{36} = \frac{1}{18}$$

* This is discrete because we know all possible outcomes. too bad we can't know everything all the time

Definition 4.1.3 Let (x, y) be a discrete bivariate random vector then the function $f(x,y)$ from $\mathbb{R}^2 \to \mathbb{R}$ defined by $P(X=x, Y=y)$ is called a joint probability mass function, or joint pmf, of (X,Y). If it is necessary to stress the fact that f is the joint pmf of the vector (X,Y) rather than some other vector function we use $f_{X,Y}(x,y)$.

Let's continue the previous example:

Y \ X	2	3	4	5	6	7	8	9	10	11	12
0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$
1	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0
2	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0
3	0	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0	0
4	0	0	0	0	$\frac{1}{18}$	0	$\frac{1}{13}$	0	0	0	0
5	0	0	0	0	0	$\frac{1}{18}$	0	0	0	0	0

↑ This would be painstaking to think about. It's simple, but it would take ages of consideration.

- Now let's talk probabilities : $P(X,Y) = \Sigma f(x,y)$

ie: $P(X=7, Y\leq 4) = f(7,4) + f(7,3) + f(7,2) + f(7,1) + f(7,0)$

$= \sum_{i=0}^{4} f(7,i)$

$= 0 + \frac{1}{18} + 0 + \frac{1}{18} = \frac{2}{18} = \frac{1}{9}$

- Expected Value: Let $g(x,y)$ be a real valued function defined on all possible (x,y) of the discrete random vector (X,Y) then $g(x,y)$ is a RV

ie: Functions of RV are RVs

$$E_{(X,Y)\in R}(g(x,y)) = \sum_{(x,y)\in R} g(x,y) f(x,y)$$

Example: (X,Y) w/ $f_{XY}(x,y)$ = the dice example, prior
Let's find the expected value of $g(x,y) = xy$

$E[g(x,y)] = E(XY) = \sum_{(x,y)} (xy) P(X=x, Y=y)$ * This would be very annoying to calculate
$= 13\frac{11}{18}$

Expected Value Properties

* Note these properties are the same as the univariate case

$$E(ag_1(X,Y) + bg_2(X,Y) + c) = aE(g_1(X,Y)) + bE(g_2(X,Y)) + c$$

PMF Properties

1) $f_{X,Y}(x,y) \geq 0 \quad \forall (x,y) \in \mathbb{R}^2$
2) $\sum_{x,y \in \mathbb{R}} f_{XY}(x,y) = 1$

Example

Define $f(x,y)$ by:

		y = 0	1
x	0	$\frac{1}{6}$	$\frac{1}{6}$
	1	$\frac{1}{3}$	$\frac{1}{3}$

0 otherwise

$f(x,y)$ is a ledgitament pmf b/c
1) $f(x,y) \geq 0 \quad \forall (x,y) \in \mathbb{R}^2$
2) $\sum_{(x,y)\in\mathbb{R}^2} f(x,y) = \frac{1}{6} + \frac{1}{6} + \frac{1}{3} + \frac{1}{3}$ + SO MANY 0s $= 1$

Theorem 4.1.6

(X,Y) is a bivariate RV w/ joint pmf $f_{X,Y}(x,y) = P(X=x, Y=y)$ then

- Marginal of X = $f_X(x) = P(X=x) = \sum_{y\in\mathbb{R}} f_{X,Y}(x,y)$
- Marginal of Y = $f_Y(y) = P(Y=y) = \sum_{x\in\mathbb{R}} f_{X,Y}(x,y)$

Proof: $f_X(x) = P(X=x)$
$= P(X=x, -\infty < Y < \infty)$ * This makes sense as $P(-\infty < Y < \infty) = 1$
$= P((X,Y) \in \{(x,y) \mid X=x, -\infty < Y < \infty\})$
$= \sum_{(x,y)\in\mathbb{R}} f_{X,Y}(x,y)$ * Def of multivariate pmf
$= \sum_{y\in\mathbb{R}} f_{X,Y}(x,y)$ * You may not buy this right away but I promise the example clears things up

$f_Y(y)$ follows same logic

* idea the pmf $f_X(x) = P(X=x)$
$= P(X=x \text{ and } Y = \text{anything})$ } this makes sense as observations are paired

Example 4.1.7 Use, not the previous example, but the example before that; you know the one w/ the ridiculous pmf table?

this is clever stuff!!

Anyways, say we want to find the marginal of Y, $f_Y(y)$. We can use our fancy new theorem to say this:

$$f_Y(0) = \sum_{x \in \mathbb{R}} f_{X,Y}(x,y) = f_{XY}(2,0) + f_{XY}(4,0) + f_{XY}(6,0) + f_{XY}(8,0) + f_{XY}(10,0) + f_{X,Y}(12,0) = \frac{1}{6}$$

* So we see that Y is fixed and we sum all outcomes $\ni Y=y=0$. The idea is that $P(Y=y) = P(X=\text{anything}, Y=y)$ ie Prob $Y=y$ is the same as all bivariate outcomes w/ $Y=y$

- Now we can do fun things like: $P(Y \leq 2) = \sum_{i=0}^{2} F_Y(i) = f_Y(0) + f_Y(1) + f_Y(2) = \frac{2}{3}$
- Don't forget $E(Y^3) = \sum_y y^3 f_Y(y) = 0^3 f_Y(0) + 1^3 f_Y(1) + \dots + 5^3 f_Y(5) = 20\frac{11}{18}$

* Marginal Distributions **DO NOT** determine joint distribution *

↑ This means it's important!

↑ I can do an example but you can just look at pg. 144!

CONTINUOUS Well, let's do the same thing all over again but now we don't know everything, there are infinite outcomes Ahhh!!!

Theorem 4.1.10 A function $f_{XY}(x,y)$ from $\mathbb{R}^2 \to \mathbb{R}$ is called a joint probability density function or joint pdf of the continuous bivariate random vector (X,Y) if $\forall (X,Y) \in \mathbb{R}^2$ in the sample space (A)

$$P((X,Y) \in A) = \iint_A f(x,y)\,dx\,dy$$

* This isn't too scary. It's just like the univariate case but in 3D ahh! 3D!

Expectations Also: $E(g(x,y)) = \iint_{\mathbb{R}^2} g(x,y) f(x,y)\,dx\,dy$ * Note $\mathbb{R}^2$ we really integrate over our support $\forall (X,Y) \in A$ or $\mathbb{R}^2$ w/ an indicator func in $f_{XY}(x,y)$

Marginals

We should probably mention the way to get the marginal distributions in the continuous case

$f_X(x) = \int_{\mathbb{R}} f_{XY}(x,y)\,dy \qquad f_Y(y) = \int_{\mathbb{R}} f_{XY}(x,y)\,dy$

Joint PDF Properties

1) $f_{XY}(x,y) \geq 0 \quad \forall\ (x,y) \in A$
2) $\iint_{\mathbb{R}} f_{XY}(x,y) = 1$

Example: Let $f_{X,Y}(x,y) = \begin{cases} 6xy^2 & \overbrace{0<x<1,\ 0<y<1}^{A} \\ 0 & \end{cases}$

Let's make sure this is valid

① $6xy^2 \geq 0 \quad \forall\ 0<x,y<1$

② $\int_0^1\int_0^1 6xy^2\,dx\,dy = \int_0^1 3x^2y^2\Big|_0^1 dy = \int_0^1 3y^2\,dy = y^3\Big|_0^1 = 1$

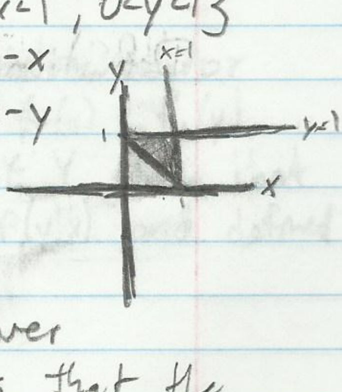
$\left.\begin{array}{l} f_X(x) = \int_0^1 6xy^2\,dy = 2xy^3\Big|_0^1 = 2x \\ f_Y(x) = \int_0^1 6xy^2\,dx = 3x^2y^2\Big|_0^1 = 3y^2 \end{array}\right\}$ marginals

Let's consider $P(X+Y \geq 1)$ we can integrate over the new set of interest $B = \{(x,y) \mid x+y \geq 1,\ 0<x<1,\ 0<y<1\}$

$x+y \geq 1 \rightarrow y \geq 1-x$ OR $\rightarrow x \geq 1-y$

So $P(X+Y \geq 1) = \int_0^1\int_{1-x}^1 6xy^2\,dy\,dx = \frac{9}{10}$

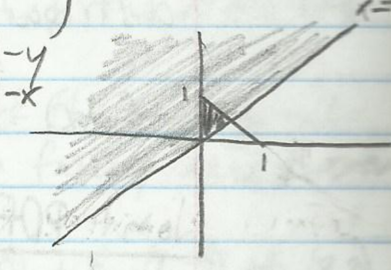
$= \int_0^1\int_{1-y}^1 6xy^2\,dy\,dx = \frac{9}{10}$



* We note that the area we integrate over is only half the support - this tells us that the 3d distribution has a larger height over this region

<u>Example</u> Let $f_{x,y}(x,y) = \begin{cases} e^{-y} & 0 < x < y < \infty \\ 0 & \end{cases}$

Let's try again to find $P(X+Y \geq 1)$ $\quad x \geq 1-y$, $\; y \geq 1-x$



- Since $X+Y \geq 1$ is unbounded we will look at it's bounded counterpart $X+Y<1$

$$
\begin{aligned}
P(X+Y \geq 1) &= 1 - P(X+Y<1) \\
&= 1 - \int_0^{1/2}\int_x^{1-x} e^{-y}\, dy\, dx \\
&= 1 + \int_0^{1/2}\int_x^{1-x} e^{u}\, du\, dx \\
&= 1 + \int_0^{1/2} e^{-y}\Big|_x^{1-x} dx \\
&= 1 + \int_0^{1/2} e^{-(1-x)} - e^{-x}\, dx \\
&= 1 + \int_0^{1/2} e^{-1+x} - \int e^{-x}\, dx \\
&= 1 + e^{-1+x}\Big|_0^{1/2} + e^{-x}\Big|_0^{1/2} \\
&= 1 + \left(\tfrac{1}{\sqrt{e}} - \tfrac{1}{e}\right) + \left(\tfrac{1}{\sqrt{e}} - 1\right) \\
&= \tfrac{2}{\sqrt{e}} - \tfrac{1}{e}
\end{aligned}
$$

⇑ It is important to graph the support! ☺

<u>JOINT CDF</u> $\quad F_{xy}(x,y) = P(X \leq x, Y \leq y) \quad \forall (x,y) \in \mathbb{R}^2$

$\longrightarrow F_{x,y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{x,y}(x,y)\, dx\, dy$

$\ni \frac{\partial^2}{\partial x \partial y} F(x,y) = f_{x,y}(x,y)$ (via Fundamental Theorem of Calc)

<u>4.2 Conditional Dist & Indep.</u> (Multicolinearity idea)

Definition 4.2.1 Let (X,Y) be a discrete bivariate random vector w/ joint pmf $f(x,y)$ and marginal pmfs $f_X(x)$, $f_Y(y)$. $\forall x \ni P(X=x) = f_X(x) > 0$ the conditional pmf of Y given $X=x$ is

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

Analagous for x

$$f_X(x|y) = P(X=x \mid Y=y) = \frac{f_{x,y}(x,y)}{f_X(x)}$$

• To show $f_Y(y|x)$ is a valid pmf

1) $f_Y(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} \geq 0$ as the numerator and denom ≥ 0

2) $\sum_y f_Y(y|x) = \frac{\sum_y f_{XY}(x,y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1$ as $\sum_y f_{XY}(x,y) = f_X(x)$ ← marginal definition

EXAMPLE • Define the joint pmf of (X,Y) by this table

		y: 10	20	30
x	0	1/9	1/9	0
	1	1/6	2/9	1/6
	2	0	0	2/9

$f_X(0) = 1/9 + 1/9 = 2/9$

$f_X(1) = 1/6 + 2/9 + 1/6 = 5/9$

$f_X(2) = 2/9$

$\Sigma = 9/9 = 1$

$f(Y=10|X=0) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{XY}(0,10)}{f_X(0)} = \frac{1/9}{2/9} = \frac{1}{9}\cdot\frac{9}{2} = \frac{1}{2}$

$f(Y=20|X=0) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{XY}(0,20)}{f_X(0)} = \frac{1/9}{2/9} = \frac{1}{2}$

$f(Y=10|X=1) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{XY}(1,10)}{f_X(1)} = \frac{1/6}{5/9} = \frac{1}{6}\cdot\frac{9}{5} = \frac{3}{10}$

⋮ We could keep going and going but you get the idea

But, look here this is different

$P(Y>10|X=1) = f(20|1) + f(30|1) = \frac{7}{10}$

* Note we've just covered discrete conditional since it doesn't make sense to say $P(X=x)$ for X a continuous RV. This is covered in the miscellanea section future you will read. *

Def 4.2.3 Let (X,Y) be a continuous bivariate random vector w/ joint pdf $f_{XY}(x,y)$ and marginals $f_X(x)$ & $f_Y(y)$. $\forall x \ni f_X(x) > 0$ the conditional pdf of Y given that $X=x$ is the function of y denoted by $f(y|x)$ and defined by

$f(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$

Analagously for $f(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$

Example. Let (X,Y) have joint pdf $f(x,y) = e^{-y}$ $\{0<x<y<\infty\}$
We wish to find $f(y|x)$

$f_X(x) = \int_{(x,y)} f(x,y)\,dy$ — As the support is $0<x<y<\infty$ our upper bound of y is ∞, our lower bound of y is x

$= \int_x^\infty e^{-y}$

$= -e^{-y}\Big|_x^\infty = 0 - -e^{-x} = e^{-x}$ for $x>0$

$\therefore$ we see X is exponential

- Now we can find $f(y|x)$

$$f(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = e^{-y}/e^{-x} = \frac{1/e^y}{1/e^x} = \frac{1}{e^y}\cdot\frac{e^x}{1} = e^{x-y} = e^{-(y-x)}\quad y>$$

$\therefore$ we see $Y|X$ is a location-scale exponential family

– We can keep going let's find Expected values & variance

$$E(Y|X=x) = \int_x^\infty y e^{-(y-x)}\,dy = 1+x$$

$$Var(Y|X=x) = E(Y^2|x) - (E(Y|x))^2 = \int_x^\infty y^2 e^{-(y-x)}\,dy - \left(\int_x^\infty y e^{-(y-x)}\,dy\right)^2 = 1$$

- Def: By rearranging $f(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$
We get

$$f_{XY}(x,y) = f(y|x)\,f_X(x)$$

- Note $E(g(Y)|x)$ is a function of x ie it is different for diff values of
$E(g(X)|y)$ is a function of y ie it is different for diff values of y

Def 4.2.5 Let (X,Y) be a bivariate random vector w/ joint pdf or pmf $f(x,y)$ and marginal pdfs or pmfs $f_X(x)$ and $f_Y(y)$. Then X and Y are independent RV if $\forall\, x\in\mathbb{R}\ y\in\mathbb{R}$ $f_{X,Y}(x,y) = f_X(x)\,f_Y(y)$

- Note we can then rewrite • $f(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_X(x)\,f_Y(y)}{f_X(x)} = f_Y(y)$

• $\int f(y|x) = \int f_Y(y)$

Example: given $f(10,3) = \frac{1}{5}$, $f_x(10) = \frac{1}{2}$, $f_y(3) = \frac{1}{2}$

- as $\frac{1}{5} \neq \frac{1}{2} \cdot \frac{1}{2}$ ∴ not independent as that relationship must hold for all x, y

<u>Lemma 4.2.7</u> Let (X,Y) be a bivariate random vector w/ joint pdf or pmf $f(x,y)$. Then X and Y are independent RV iff $\exists$ functions $g(x)$ $h(y)$ $\ni$ $\forall$ $x \in \mathbb{R}$ $y \in \mathbb{R}$
$f(x,y) = g(x) h(y)$

example: $f_{X,Y}(x,y) = \frac{1}{384} x^2 y^4 e^{-y-(x/2)}$

- I don't want to find the marginals, do you? Let's use our friendly lemma 4.2.7

$$f_{X,Y}(x,y) = g(x)\, h(y) \quad \text{if} \quad g(x) = \frac{1}{384} x^2 e^{-x/2}, \quad h(y) = y^4 e^{-y}$$

So we can say, this easily, that x and y are independent

- It turns out that if x and y are independent things are a tad easier for us

<u>Theorem 4.2.10</u> For independent X and Y

<u>If</u>

a) $\forall$ $A, B \subset \mathbb{R}$ $P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$

b) Let $g(x)$ and $h(y)$ be functions of just x or y

<u>Then</u>

$E(g(x) h(y)) = (E(g(x)))(E(h(y)))$

Easy to show, even for me, pg 155

Example: $P(X \geq 4, Y < 3) = P(X \geq 4)\,P(Y < 3) = e^{-4}(1-e^{-3})$ as X, Y ind.

- Now say we want to find $E(X^2 Y)$

$E(X^2Y) = E(X^2)E(Y)$ as X, Y indep.

$= (Var(X) + (E(X))^2)\,E(Y) = (1+1^2)(1) = 2$

Theorem 4.2.12 Let X and Y be independent w/ MGFs $M_X(t)$ and $M_Y(t)$. Then the MGF of $Z = X+Y$ is

$$M_Z(t) = M_X(t)\,M_Y(t)$$

This is proved using def of MGF and ind. exp. value rules

Example 4.2.13 Let $X \sim n(\mu, \sigma^2)$ and $Y \sim n(\gamma, \tau^2)$ be independent

For $M_X(t) = \exp(\mu t + (\sigma^2 t^2/2))$ $M_Y(t) = \exp(\gamma t + (\tau^2 t^2)/2)$

$Z = X + Y$

$$M_Z(t) = \exp\left[\left(t(\mu + \sigma^2 t + \gamma + \tau^2 t)/2\right)\right]$$

- And since we're smart and we know MGFs uniquely describe pdf/pmf we know $Z \sim N(\mu + \gamma, \sigma^2 \tau^2)$. Pretty snazzy, huh?

4.3 Bivariate Transformation. - This will be like theorems 2.1.5 and 2.1.8 from ch. 2

- Let (X, Y) be a bivariate Random vector
- We want to consider (U, V) defined by $U = g_1(X,Y)$ & $V = g_2(X,Y)$
  - We look at the sample space in this way
    - $B \subseteq \mathbb{R}^2 \ni (U,V) \in B$ iff $(X,Y) \in A = \{(x,y) \mid (g_1(x,y), g_2(x,y)) \in B\}$
    - We ~~can~~ can completely determine the distribution of (U,V) from that of (X,Y)
    - For the discrete case:

$$f_{U,V}(u,v) = P(U=u, V=v) = P((X,Y) \in A_{u,v}) = \sum_{(x,y)\in A} f_{X,Y}(x,y)$$

$\boxed{\text{Example}}$: $X \sim \text{Poisson}(\theta)$, $Y \sim \text{Poisson}(\lambda)$ → X and Y are independent

- As X and Y are independent we can derive the joint dist

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$
$$= \left(\frac{e^{-\theta}\theta^x}{x!}\right)\left(\frac{e^{-\lambda}\lambda^y}{y!}\right) \quad \theta \geq 0,\ \lambda \geq 0 \ni \theta, \lambda \in \mathbb{Z}$$

↑ this is our support $A = \{(x,y) \mid x = 0,1,2\ldots\ y = 0,1,2\ldots\}$

- Let's now consider $U = X+Y$ and $V = Y$

i.e. $g_1(x,y) = x+y$ $\quad g_2(x,y) = y$

$g_1(x,y) = 0,1,2,3\ldots$ $\quad g_2(x,y) = 0,1,2,3\ldots$

- We now can define $B = \{(u,v) \mid v = 0,1,2,\ldots\ u = v+0, v+1, v+2, \ldots\}$

$$\therefore f_{U,V}(u,v) = f_{X,Y}(u-v,v) = \left(\frac{e^{-\theta}\theta^{(u-v)}}{(u-v)!}\right)\left(\frac{e^{-\lambda}\lambda^v}{v!}\right) \ni v = 0,1,2,\ldots \quad u = v, v+1, v+2\ldots$$

- If we wanted we can calculate the marginals

$$f_U(u) = \sum_{v=0}^{\infty} f_{U,V}(u,v) \Rightarrow \text{Note we will cap this and find } \sum_{v=0}^{u} f_{U,V}(u,v)$$

$$= \sum_{v=0}^{u}\left(\frac{e^{-(u-v)}}{(u-v)!}\right)\left(\frac{e^{-\lambda}\lambda^v}{v!}\right) \Rightarrow \text{which after some annoying algebra} = \frac{e^{-(\theta+\lambda)}(\theta+\lambda)^u}{u!} \quad \text{for } u = 0,1,2\ldots$$

notice $U = X - Y \sim$ is poisson$(\theta+\lambda)$ → pretty cool, eh?

<u>Theorem 4.3.2</u> If $X \sim \text{Poisson}(\theta)$ and $Y \sim \text{Poisson}(\lambda) \ni X, Y$ are independent then $U = (X+Y) \sim \text{Poisson}(\theta+\lambda)$

- If we assume the transformation is onto and one to one

(X,Y)
$(U,V) \ni u = g_1(x,y) \quad v = g_2(x) \rightarrow x = h_1(u,v) \quad y = h_2(u,v)$

- We can then calculate the jacobian

$$J = \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dv} \\ \frac{dy}{du} & \frac{dy}{dv} \end{vmatrix} = \left| \frac{dx}{du}\frac{dy}{dv} - \frac{dx}{dv}\frac{dy}{du} \right|$$

- If we assume J is not singular ie $J \neq 0$ we can say

$$f_{u,v}(u,v) = f_{x,y}(h_1(u,v), h_2(u,v)) \cdot |J|$$

Example $X \sim \text{beta}(\alpha, \beta)$, $Y \sim \text{beta}(\alpha+\beta, \gamma)$ $\rightarrow$ X,Y are independent $X,Y \in (0,1)$

- As x,y are independent $f_{x,y}(x,y) = \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\right]\left[\frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha+\beta)\Gamma(\gamma)}(y)^{\alpha+\beta-1}(1-y)^{\gamma-1}\right]$

$$= \left[\frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} x^{\alpha-1}\, 1-x^{\beta-1}\, y^{\alpha+\beta-1}(1-y)^{\gamma-1}\right]$$

- Consider $U = XY \quad V = X$

$u \in (0,v) \quad v \in (0,1)$

$A = \{(X,Y) \mid X, Y \in (0,1)\} \quad B = \{(U,V) \mid 0 < u < v < 1\}$

- We can solve $X = V \quad Y = \frac{U}{V} \ni$ one to one + onto

$$J = \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dv} \\ \frac{dy}{du} & \frac{dy}{dv} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ \frac{1}{v} & \frac{-u}{2v^2} \end{vmatrix} = \frac{-1}{v}$$

Thus $f_{u,v}(u,v) = f_{x,y}\left(v, \frac{u}{v}\right)\left|\frac{-1}{v}\right|$

$$= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} v^{\alpha-1}(1-v)^{\beta-1}\left(\frac{u}{v}\right)^{\alpha+\beta-1}\left(1-\frac{u}{v}\right)^{\gamma-1}\left|\frac{-1}{v}\right|$$

on $0 < u < v < 1$

We can then find the marginals!

$$f_U(u) = \int_u^1 f_{U,V}(u,v)\,dv = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1} \int_u^1 \left(\frac{u}{v}-u\right)^{\beta-1}\left(1-\frac{u}{v}\right)^{\gamma-1}\frac{u}{v^2}\,dv$$

- If we're clever, I'm not sure anyone is this clever, we can let $y = \left(\frac{u}{v}-u\right)/(1-u)$ $\quad dy = \frac{-u}{x^2(1-u)}\,dv$ which yields

$$f_U(u) = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} \cdot u^{\alpha-1}(1-u)^{\beta+\gamma-1} \underbrace{\int_0^1 y^{\beta-1}(1-y)^{\gamma-1}\,dy}_{\text{see this and think GAMMA!}}$$

$$\therefore f_U(u) = \frac{\Gamma(\alpha+\beta+\gamma)\cancel{\Gamma(\beta)\Gamma(\gamma)}}{\Gamma(\alpha)\cancel{\Gamma(\beta)\Gamma(\gamma)}\Gamma(\beta+\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1}$$

$$= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\gamma)\Gamma(\beta+\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1}$$

$\therefore U \sim \text{beta}(\alpha, \beta+\gamma)$

↑ this is a ton of work but an interesting result, no?

<u>Example</u> $\left.\begin{array}{l} X \sim N(0,1) \\ Y \sim N(0,1)\end{array}\right\} \to$ X, Y are independent $\quad -\infty < X, Y < \infty \Rightarrow A = \mathbb{R}^2$

$U = X+Y \qquad V = X-Y \quad \longrightarrow \quad g_1(x,y) = x+y \quad g_2(x,y) = x-y$

$Y = \frac{U-V}{2} \qquad X = \frac{U+V}{2} \quad \longrightarrow \quad h_1(u,v) = \frac{u-v}{2} \quad h_2(x,y) = \frac{u+v}{2}$

As x, y are independent $f_{X,Y}(x,y) = \frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}}$

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{-1}{2} \end{vmatrix} = -\frac{1}{2}$$

$$\therefore f_{U,V}(u,v) = f_{X,Y}\left(\frac{u-v}{2}, \frac{u+v}{2}\right)\left|\frac{-1}{2}\right| = \frac{1}{4\pi} e^{-\left(\frac{\left(\frac{u-v}{2}\right)^2 + \left(\frac{u+v}{2}\right)^2}{2}\right)} = \frac{1}{4\pi} e^{-\left(\frac{(u-v)^2+(u+v)^2}{8}\right)}$$

ASIDE → $(u-v)^2 + (u+v)^2 = u^2 - 2uv + v^2 + u^2 + 2uv + v^2 = 2(u^2+v^2)$

$$= \frac{1}{4\pi} e^{-(u^2+v^2)/4}$$

- Are U and V independent?

$$= \Big(\underset{\uparrow\, g(u)}{\frac{1}{4\pi} e^{-\frac{u^2}{4}}}\Big)\Big(\underset{\uparrow\, h(v)}{e^{-v^2/4}}\Big)$$

$\therefore$ independent

* Look how clever we are! We used our Lemma instead of finding the marginals and showing $f_{UV}(u,v) = f_U(u) f_V(v)$

Theorem 4.3.5: If X and Y are independent RV and $g(x)$ is a function of only x and $h(y)$ is a function of only y then RVs $U=g(x)$ & $V=h(y)$ are independent

Proof: $F_{U,V}(u,v) = P(U \le u, V \le v)$

same statement in terms of x, y $\rightarrow$ $= P(X \in A_u, Y \in B_v)$ $: A = \{x \mid g(x) \le u\}$ $B = \{x \mid h(y) \le v\}$

$= P(X \in A_u)\, P(Y \in B_v)$: as x, y are independent

Then $f_{UV}(u,v) = \frac{\partial^2}{\partial u \partial v} F_{UV}(u,v) = \frac{d}{du} P(X \in A_u) \frac{d}{dv} P(Y \in B_v)$

$\therefore$ By the same Lemma U and V are independent

• Note: if we are <u>just interested in one variable, say $U = X+Y$</u> we'll play match maker and find a dummy $g_2(x,y) \ni$ the bivariate transformation is one to one and onto, continue on as usual

Q• WHAT IF THERE'S A FLAW IN THE MATRIX? WHAT IF WE AREN'T IN THE LAND OF ONE TO ONE and onto?

A• We will, like in theorem 2.1.8, partition the support

• Suppose $A_0, A_1, \ldots A_K$ form a partition of $A \ni (X,Y) \rightarrow (U,V)$ is one to one and onto $\ni A_0$ may be empty and $P((X,Y) \in A_0) = 0$

• Transform: $U = g_1(x)$ $V = g_2(x)$

$x = h_{1i}(u,v)$ $y = h_{2i}(u,v)$ $\forall\, i = 0,1,2 \ldots k$ (ie $\forall A_i$)

$$J_i = \begin{vmatrix} \frac{dx_i}{du} & \frac{dx_i}{dv} \\ \frac{dy_i}{du} & \frac{dy_i}{dv} \end{vmatrix}$$ $\Leftarrow$ Lucky us we get to calculate a jacobian $\forall$ partition of A

<u>THEN, FINALLY</u>

$$f_{U,V}(u,v) = \sum_i^k f_{XY}(h_{1i}(u,v), h_{2i}(u,v)) |J_i|$$

Example: $X \sim N(0,1)$, $Y \sim N(0,1)$ → independent $\quad -\infty < x < \infty \quad -\infty < y < \infty$

$U = \frac{X}{Y} \quad V = |Y| \quad -\infty < u < \infty \quad 0 < v < \infty$

$A_0 = \{\emptyset\}$

$A_1 = \{(u,y) \mid -\infty < u < \infty,\ -\infty < y < 0\} \quad x_1 = U(-V) \quad y_1 = -V$

$A_2 = \{(u,y) \mid -\infty < u < \infty,\ 0 < y < \infty\} \quad x_2 = UV \quad y_2 = V$

$$J_1 = \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dv} \\ \frac{dy}{du} & \frac{dy}{dv} \end{vmatrix} = -v(-1) - (-u)(0) = \boxed{v}$$

$$J_2 = v(1) - (u)(0) = \boxed{v}$$

$$f_{U,V}(u,v) = \sum_i^2 f_{X,Y}(x_i, y_i)|J_i| \quad \text{where } f_{X,Y}(x,y) = \frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}}$$

$$= \frac{1}{2\pi} e^{-\frac{((-uv)^2 + (-v)^2)}{2}} |-v| + \frac{1}{2\pi} e^{-\frac{((uv)^2+v^2)}{2}} |v|$$

$$= \frac{|v|}{\pi} e^{-\frac{((uv)^2+v^2)}{2}} \quad \text{for } -\infty < u < \infty \ \&\ 0 < v < \infty$$

I should probably write this over here ← Marginals $\quad f_U(u) = \int_0^\infty \frac{|v|}{\pi} e^{-\frac{((uv)^2+v^2)}{2}} dv$

$$= \frac{1}{\pi(u^2+1)} \quad -\infty < u < \infty \quad \text{(Cauchy)}$$

4.4 Hierarchical Models - here we can have models of models!

Example $\quad X|Y \sim \text{binomial}(Y, p)$ ← Y is fixed here

$Y \sim \text{Poisson}(\lambda)$

This can be thought of as follows # of eggs layed by insect is modeled w/ Poisson, the number that survive by a binomial

ie Y = # of eggs $\quad X$ = # that survive

- Now that we have the idea let's do some math

$$P(X=x) = \sum_{y=0}^{\infty} P(X=x, Y=y)$$

$$= \sum_{y=0}^{\infty} P(X=x|Y=y)\,P(Y=y)$$ * as $f_{X,Y}(x,y) = f_X(x|y)\,f_Y(y)$

$$= \sum_{y=x}^{\infty} P(X=x|Y=y)\,P(Y=y)$$ * as y is at least x as # of eggs $\geq$ # survived

$$= \sum_{y=x}^{\infty}\left[\binom{y}{x} p^x (1-p)^{y-x}\right]\left[\frac{e^{-\lambda}\lambda^y}{y!}\right]$$ * Remember $X|Y \sim \text{binomial}(Y,p)$, $Y \sim \text{poisson}(\lambda)$

$$= \sum_{y} \frac{y!}{x!(y-x)!} p^x (1-p)^{y-x} (e^{-\lambda}\lambda^y / y!)$$

$$= \frac{p^x e^{-\lambda}}{x!(1-p)^x} \sum_{y=x}^{\infty} ((1-p)^y \lambda^y)/(y-x)!$$ which we can show as a poisson kernel for dummy $z = y-x$

$$= \frac{p^x e^{-\lambda}}{x!(1-p)^x} \sum_{t=0}^{\infty}\left[((1-p)\lambda)^{t+x} / t!\right]$$

$$= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{t=0}^{\infty}\left[((1-p)\lambda)^t / t!\right] = \frac{\lambda p^x e^{-\lambda}}{x!}\left(e^{(1-p)\lambda}\right)$$

$$= \frac{(\lambda p)^x}{x!} e^{-\lambda p}$$ * So after all of this we now know $X \sim \text{Poisson}(\lambda p)$

So $E(X) = \lambda p$
$\quad var(X) = \lambda p$

Theorem 4.4.3 If X and Y are any two RVs then $E(X) = E(E(X|Y))$

$$E(X) = \iint x f_{XY}(x,y)\,dx\,dy = \int\left[\int x f(x|y)\,dx\right] f_Y(y)\,dx$$
$$= E(E(X|Y))$$

Example $E(X) = E(E(X|Y)) = E(pY)$ * as $X|Y \sim$ bin
$= pE(Y)$ * exp val rules
$= p\lambda$ * as $Y \sim$ poisson

Def 4.4.4. A RV X is said to have a mixture distribution if the distribution of x depends on a quantity that also has a distribution

Note: In the previous example $X \sim$ Poisson (λp) distribution is a mixture as it is the result of combining binomial (Y, p) with $Y \sim$ poisson (λ).

Example: $X|Y \sim$ binomial (Y, p)
$Y|\Lambda \sim$ poisson (Λ)
$\Lambda \sim$ exponential (β)

$$\begin{aligned} E(X) &= E(E(X|Y)) \\ &= E(pY) \qquad \text{as above} \\ &= E(E(pY|\Lambda)) \\ &= E(p\Lambda) \\ &= p\beta \end{aligned}$$

↑ Note this is easier than finding transformations, marginals and expectations the hard way

ie: we could have found the pdf of y then the pdf of x and after all that work we could take, a probably disgusting, integral but we'll spare ourselves of that

Example: $Y|\Lambda \sim \text{Poisson}(\Lambda)$
$\Lambda \sim \text{gamma}(\alpha, \beta)$
$\therefore Y \sim$ neg. binomial

Example: $X|K \sim \chi^2_{p+2K}$ To show noncentral chi-squ. dist
$K \sim \text{Poisson}(\lambda)$

$$f(x|\lambda, p) = \sum_0^\infty \left[\frac{x^{p/2+k-1} e^{-x/2}}{\Gamma(p/2+k)\, 2^{p/2+k}} \cdot \frac{\lambda^k e^{-\lambda}}{k!} \right]$$

$$\begin{aligned} E(X) &= E(E(X|K)) \\ &= E(p+2K) \\ &= p + 2\lambda \end{aligned}$$

Example $X|P \sim \text{binomial}(P)$ $i = 1, 2, 3, \ldots, n$
$P \sim \text{beta}(\alpha, \beta)$

$$E(X) = E(E(X|P)) = E(nP) = n\left(\frac{\alpha}{\alpha+\beta}\right)$$

Theorem 4.4.7 $\forall$ 2 RV X, Y

$$Var(X) = E(Var(X|Y)) + Var(E(X|Y))$$ provided the expectations exist

Proof $Var(X) = E((X - E(X))^2) = E([X - E(X|Y) + E(X|Y) - E(X)]^2)$
complete the square

$$\begin{aligned} &= E((X - E(X|Y))^2) + E((E(X|Y) - E(X))^2) \\ &\quad + 2E((X - E(X|Y))(E(X|Y) - E(X))) \\ &= E((X - E(X|Y))^2) + E((E(X|Y) - E(X))^2) \end{aligned}$$

$\vdots$

p. 168

Example $Var(x) = Var(E(x|p)) + E(Var(X|p))$

- Let $E(X|p) = nP$ and $P \sim beta(\alpha, \beta)$
- $Var(E(X|p)) = var(nP) = n^2 \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Also since $X|P \sim binomial(n, P)$, $Var(x|p) = nP(1-p)$ then

$$E(var(x|p)) = n\,E(P(1-P)) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \underbrace{p(1-p)p^{\alpha-1}(1-p)^{\beta-1}\,dp}_{\text{beta kernal}}$$

$$E(Var(x|P)) = n \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left[\frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}\right] = n\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}$$

Thus

$$var(x) = E(var(x|p)) + var(E(x|p))$$
$$= \frac{n\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)} + \frac{n^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

4.5 Cov and corr - measures of the strength of a *LINEAR* relationship

ie X = weight of water sample, Y = volume of water sample } clearly there will be a strong relationship

ie X = height of person, Y = weight of person } There will be a relationship but maybe not so strong

ie X = weight of a person, Y = volume of a water sample } No relation

Notation
$\mu_X = E(X)$
$\mu_Y = E(Y)$
$\sigma_X^2 = Var(X) \in (0, \infty)$
$\sigma_Y^2 = Var(Y) \in (0, \infty)$

Covariance of X, Y $\quad Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y))$

Correlation of X, Y $\quad \rho_{XY} = \dfrac{COV(X,Y)}{\sigma_X \sigma_Y}$
(or correlation coefficient)

① • If large values of X tend to be observed w/ large Y and small x with small y then $Cov(x,y) > 0$

• If $X > \mu_X$ then $Y > \mu_Y$ is likely to be true and the product $(X-\mu_X)(Y-\mu_Y)$ will be positive. DUH
• If $X < \mu_X$ then $Y < \mu_Y$ is likely to be true and the product $(X-\mu_X)(Y-\mu_Y)$ will be positive
• Thus $Cov(X,Y) = E((X-\mu_X)(Y-\mu_X)) > 0$

② If large values of X tend to be observed w/ small values of y and small x w/ large y then $cov(x,y) < 0$ by similar logic above

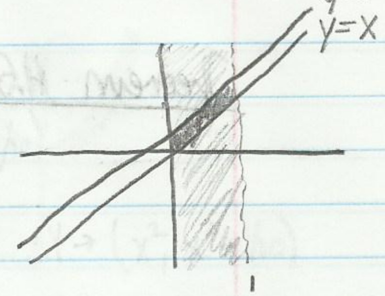
• Cov can be any number. It will tell us what we've defined above but doesn't say anything about the strength of the relationship. For this, we will use the correlation which is between -1 and 1

$\rho_{xy} = 1 \rightarrow$ perfect positive relationship
$\rho_{xy} = -1 \rightarrow$ perfect negative relationship

Theorem 4.5.3 - For RVs $X, Y \rightarrow Cov(X,Y) = E(XY) - \mu_X\mu_Y$

Proof
$$\begin{aligned} Cov(X,Y) &= E((X-\mu_X)(Y-\mu_Y)) = E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y) \\ &= E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X\mu_Y \\ &= E(XY) - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y \\ &= E(XY) - \mu_Y\mu_X \end{aligned}$$

Example $f_{X,Y}(x,y) = 1$ on $0 < x < 1$, $x < y < x+1$



$f_X(x) = \int_x^{x+1} 1\, dy$
$y\Big|_x^{x+1}$
$x+1-x = 1$

$f_Y(y) = \int_0^1 1\, dx$
$= y\Big|_0^1$
$= 1$

The book gives: $\mu_x = \frac{1}{2}$ $\sigma_x^2 = \frac{1}{12}$ $f_X(x) = 1\ I(0<x<1)$

$\mu_y = 1$ $\sigma_y^2 = \frac{1}{6}$ $f_Y(x) = \begin{cases} y & 0 < y < 1 \\ 2-y & 1 \le y < 2 \end{cases}$

$$E(XY) = \int_0^1 \int_x^{x+1} xy(1)\, dy\, dx = \int_0^1 \frac{1}{2} xy^2 \Big|_x^{x+1} = \int_0^1 \frac{1}{2}x(x+1)^2 - \frac{1}{2}x^3\, dx$$
$$= \int_0^1 \frac{1}{2}x(x^2+2x+1) - \frac{1}{2}x^3\, dx = \int_0^1 \frac{1}{2}x^3 + x^2 + \frac{1}{2}x - \frac{1}{2}x^3\, dx = \int_0^1 x^2 + \frac{1}{2}x\, dx$$
$$= \frac{1}{3}x^3 + \frac{1}{4}x^2 \Big|_0^1 = \frac{1}{3} + \frac{1}{4} = \frac{4+3}{12} = 7/12$$

$$\text{Cov}(X,Y) = E(XY) - \mu_x\mu_y = \frac{7}{12} - \left(\frac{1}{2}\right)(1) = \frac{7}{12} - \frac{6}{12} = \boxed{1/12}$$

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{1/12}{\sqrt{\frac{1}{12}}\sqrt{\frac{1}{6}}} = \frac{1/12}{1/\sqrt{72}} = \frac{1/12}{1/6\sqrt{2}} = \frac{1}{12} \cdot \frac{6\sqrt{2}}{1} = \frac{\sqrt{2}}{2} = \boxed{\frac{1}{\sqrt{2}}}$$

Theorem 4.5.5 If X and Y are random variables then $\text{Cov}(X,Y) = 0$ and $\rho_{XY} = 0$

Proof $\text{Cov}(X,Y) = E(XY) - \mu_X\mu_Y \Rightarrow E(XY) = E(X)E(Y)$ when ind
$= 0$

thus $\rho_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = 0$

Theorem 4.5.6
- If X and Y are any two random variables and a and b are any two constants then $\text{Var}(aX+bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\,\text{cov}(X,Y)$
- If X and Y are independent $\text{Var}(aX+bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$

Proof $E(aX+bY) = a\mu_x + b\mu_y$

$$\text{Var}(aX+bY) = E\left(\left((aX+bY) - (a\mu_x + b\mu_y)\right)^2\right) = E\left(\left(a(X-\mu_x) - b(Y-\mu_y)\right)^2\right)$$
$$= E\left(a^2(X-\mu_x)^2 + b^2(Y-\mu_y)^2 - 2ab(X-\mu_x)(Y-\mu_y)\right)$$
$$= a^2E\left((X-\mu_x)^2\right) + b^2E\left((Y-\mu_y)^2\right) - 2ab\,E\left((X-\mu_x)(Y-\mu_y)\right)$$
$$= a^2\text{Var}(X) + b^2\text{Var}(Y) - 2ab\,\text{Cov}(X,Y)$$

Note: when x and y are independent $\text{Cov}(X,Y) = 0$

Note: if cov < 0 adds variation cov > 0 removes variation

Theorem 4.5.7 For any random variables X and Y

a) $-1 \le \rho_{XY} \le 1$

b) $|\rho_{XY}| = 1$ iff $\exists$ $a \ne 0$ and $b \ni P(Y = aX + b) = 1$

if $a > 0$ $\rho_{XY} = 1$, $a < 0$ $\rho_{XY} = -1$

There is a confusing proof on pg 172-173

Example $X \sim \text{Uniform}(0,1)$ $f_X(x) = 1\, I(0<x<1)$
$Z \sim \text{Uniform}(0, \frac{1}{10})$ $f_Z(z) = 10\, I(0<x<\frac{1}{10})$ > Independent
$Y = X + Z$ $(0, \frac{11}{10})$

$f_{X,Z}(x,z) = f_X(x) f(z) = (1)(10) = 10$ as x, z are ind.

Let W=X be a dummy variable (0,1)

$X = W$ $Z = Y - W$

$$J = \begin{vmatrix} \frac{dx}{dw} & \frac{dx}{dy} \\ \frac{dz}{dw} & \frac{dz}{dy} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = |1|$$

$f_{Y,X} = f_{Y,W} = f_{X,Z}(x+z, x)\, |J| = 10|1| = 10$ $I(x<y<x+\frac{1}{10};\ 0<x<1)$

$f_Y(y) = \int_0^1 10\, dw = 10$

?? $f_{Y|X} = \frac{f_{Y,X}(y,x)}{f_X(x)} = \frac{10}{1} = 10$

$y|x \sim \text{uniform}(\cancel{0}, \frac{1}{10})$

Book has

$\Downarrow$

$y|x \sim \text{uniform}(x^2, x^2 + \frac{1}{10})$

- $E(X) = \frac{1}{2}$, $E(Y) = E(X+Z) = E(X) + E(Z) = \frac{1}{2} + \frac{1}{20} = \frac{11}{20}$
- $\text{Cov}(X,Y) = E(XY) - E(X)E(Y) = E(X(X+Z)) - E(X)E(Y) = E(X^2) + E(XZ) - E(X)E(X+Z)$
$= E(X^2) + E(X)E(Z) - (E(X))^2 - E(X)E(Z) = E(X^2) - (E(X))^2$
$= \sigma_X^2 = \frac{1}{12}$
- Theorem 4.5.6 $\rightarrow$ $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) = \frac{1}{12} + \frac{1}{1200}$

$$\rho_{XY} = \frac{\frac{1}{12}}{\sqrt{\frac{1}{12}}\sqrt{\frac{1}{12} + \frac{1}{1200}}} = \sqrt{\frac{100}{101}}$$

help!, **Example** $X \sim (-1,1)$ $\Big\}$ independent $\quad f_X(x) = \frac{1}{2}\, I(-1<x<1)$

$Z \sim (0, \frac{1}{10})$ $\quad f_Z(z) = 10\, I(0<z<1/10)$

$Y = X^2 + Z \qquad W = X^2 \leftarrow$ Dummy $\qquad W \to (0,1) \qquad Y \to (x^2, x^2 + 1/10)$

$X = \pm\sqrt{W} \qquad Z = Y - W$

$$f_{X,Z} = \left(\tfrac{1}{2}\right)(10) = 5\; I(-1<x<1;\ 0<z<1/10)$$

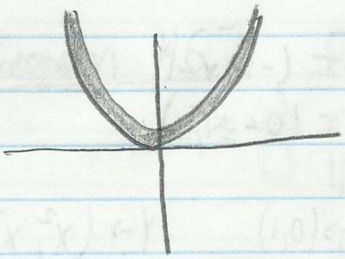
We want $f_{Y|X=x} = \dfrac{f_{X,Y}(x,y)}{f_X(x)}$

$A_0 = \{\}$, $A_1 = \{-1<x<0;\ 0<z<1/10\}$, $A_2 = \{0<x<1,\ 0<z<1/10\}$

$$A_1: J_1 = \begin{vmatrix} \frac{dx}{dw} & \frac{dx}{dy} \\ \frac{dz}{dw} & \frac{dz}{dy} \end{vmatrix} = \begin{vmatrix} \frac{1}{2\sqrt{w}} & 0 \\ -1 & 1 \end{vmatrix} = \frac{1}{2\sqrt{w}} \qquad A_2: J_2 = \begin{vmatrix} \frac{-1}{2\sqrt{w}} & 0 \\ -1 & 1 \end{vmatrix} = \frac{-1}{2\sqrt{w}}$$

$$f_{Y,W}(y,w) = \left|\frac{1}{2\sqrt{w}}\right| f_{X,Z}(\sqrt{w}, y-w) + \left|\frac{-1}{2\sqrt{w}}\right| f_{X,Z}(-\sqrt{w}, y-w) = \left|\frac{1}{2\sqrt{w}}\right| 5 + \left|\frac{1}{2\sqrt{w}}\right| 5 = \left(5/\sqrt{w}\right)$$

$$f_Y = \int_0^1 5/\sqrt{w}\, dw \qquad\qquad f_W = \int_x^{x+1/10} 5/\sqrt{w}\, dy$$



Strong relationship

$\rho_{xy} = cov(x,y) = 0$ as both measure a LINEAR relationship between x and y

Definition 4.5.10

$-\infty < \mu_x < \infty$ $\quad -\infty < \mu_y < \infty$

$0 < \sigma_x < \infty$ $\quad 0 < \sigma_y < \infty$

$-1 < \rho < 1$

Bivariate Normal

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot e^{\frac{-1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)}$$

↑ for $-\infty < x < \infty$ and $-\infty < y < \infty$

1) $f_X(x)$ is $N(\mu_x, \sigma_x^2)$
2) $f_Y(y)$ is $N(\mu_y, \sigma_y^2)$
3) ρ_{xy} is ρ
4) $aX + bY \sim N(a\mu_x + b\mu_y, a^2\sigma_x + b^2\sigma_y + 2ab\rho\sigma_x\sigma_y)$

4.6 Multivariate Distributions

PDF/PMF • $\mathbf{X} = (x_1, x_2, x_3, x_4, \ldots x_n) \subseteq \mathbb{R}^n$

$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{x}}(x_1, x_2, x_3, \ldots x_n) = P(X_1 = x, X_2 = x_2, \ldots X_n = x_n)$

$\exists \iiint_{x \in A} \cdots \int f_{\mathbf{x}}(\mathbf{x})\, dx_1\, dx_2\, dx_3 \cdots dx_n$

Expected Value • Let $g(\mathbf{x}) = g(x_1, x_2, \ldots x_n)$

CONTINUOUS

$E(g(\mathbf{x})) = \iiint \cdots \int g(x) f_X(x)\, dx_1\, dx_2 \ldots dx_n$

DISCRETE

$E(g(x)) = \sum_{x \in \mathbb{R}^n} g(x) f(x)$

Marginals • $\mathbf{X} = (x_1, x_2, x_3, \ldots x_n)$ $\quad \mathbf{X}_{sub} = (x_1, x_2, x_3, \ldots x_k)$ ← this subset is a marginal

CONTINUOUS

$f_{X_{sub}}(x_1, x_2, x_3, \ldots x_k) = \iint \cdots \int f_X(x_1, x_2, x_3, \ldots x_n)\, dx_{k+1}\, dx_{k+2} \cdots dx_n$

Discrete

$f_{X_{sub}}(x_1, x_2, x_3, \ldots x_k) = \sum_{x_{k+1}} \sum_{x_{k+2}} \sum_{x_{k+3}} \cdots \sum_{x_n} f(x_1, x_2, x_3, \ldots x_n)$

Example 4.6.1 Let $n=4$

$$f_X(x_1,x_2,x_3,x_4) = \frac{3}{4}\left(x_1^2+x_2^2+x_3^2+x_4^2\right) \qquad 0<x_i<1$$

- $\int_0^1\int_0^1\int_0^1\int_0^1 \frac{3}{4}(x_1^2+x_2^2+x_3^2+x_4^2)\,dx_1dx_2dx_3dx_4 = 1$ after much elementary calculus

- $P(X_1<\frac{1}{2},\ X_2<\frac{3}{4},\ X_4>\frac{1}{2}) = \int_{1/2}^{1}\int_0^1\int_0^{3/4}\int_0^{1/2} \frac{3}{4}(x_1^2+x_2^2+x_3^2+x_4^2)\,d_1d_2d_3d_4 = \frac{151}{1024}$

- Find marginal pdf of (X_1,X_2)

$$f_{X_1X_2}(x_1,x_2) = \int_0^1\int_0^1 \frac{3}{4}(x_1^2+x_2^2+x_3^2+x_4^2)\,dx_3dx_4$$
$$= \frac{3}{4}(x_1^2+x_2^2)+\frac{1}{2}$$

- $E(X_1X_2) = \int_0^1\int_0^1 x_1x_2\left(\frac{3}{4}(x_1^2+x_2^2)+\frac{1}{2}\right)dx_1dx_2 = \frac{5}{16}$

- $$\boxed{f_{(x_{k+1},\dots x_n)|x_1,x_2,\dots x_k)} = \frac{f(x_1,\dots x_n)}{f(x_1,\dots x_k)}}$$ You missed this in note taking

$$f_{X_3X_4|X_1X_2} = \frac{\frac{3}{4}(x_1^2+x_2^2+x_3^2+x_4^2)}{\frac{3}{4}(x_1^2+x_2^2)+\frac{1}{2}} = \frac{x_1^2+x_2^2+x_3^2+x_4^2}{x_1^2+x_2^2+\frac{2}{3}}$$

Definition 4.6.2 $n,m \in \mathbb{Z}^+$ and let $p_1,\dots,p_n$ be numbers satisfying $0\le p_i\le 1$ $i=1,\dots,n$ and $\sum_1^n p_i = 1$ then the random vector $(X_1,\dots X_n)$ has a multinomial distribution w/ m trials and cell probabilities $p_1,p_2\dots p_n$ if the joint pmf of $(X_1,\dots X_n)$ is

$$f(x_1,x_2,\dots x_n) = \frac{m!}{x_1!x_2!\cdots x_n!}\,p_1^{x_1}p_2^{x_2}\cdots p_n^{x_n} = m!\prod_{i=1}^{n}\frac{p_i^{x_i}}{x_i!}$$
$$\ni x_i>0 \text{ and } \sum_1^n x_i = m$$

m independent trials w/ n distinct outcomes

p_i is the probability of the i^{th} outcome

x_i is the # of times the i^{th} outcome occured in m trials

Example Toss a 6 sided die 10 times

$m=10,\ n=6$ $\quad p_1=\frac{1}{21},\ p_2=\frac{2}{21}\ p_3=\frac{3}{21}\ p_4=\frac{4}{21}\ p_5=\frac{5}{21}\ p_6=\frac{6}{21}$

Note $\sum_1^6 p_i = 1$

$$f(x_1,x_2,x_3,x_4,x_5,x_6)=\frac{10!}{x_1!\,x_2!\,x_3!\,x_4!\,x_5!\,x_6!}(p_1)^{x_1}(p_2)^{x_2}(p_3)^{x_3}(p_4)^{x_4}(p_5)^{x_5}(p_6)^{x_6}$$

$$f(0,0,1,2,3,4)=.0059$$

Multinomial Coefficient $\frac{m!}{x_1!\,x_2!\cdots x_n!}$ = ways to split m objects into n groups w/ x_1 in the first x_2 in the 2^{nd}

Theorem 4.6.4 Let m and $n \in \mathbb{Z}^+$. Let A be the set $x=(x_1,x_2 \ldots x_n)$ $\ni x_i \in \mathbb{Z}^+ + \{0\}$ and $\sum_1^n x_i = m$ then for $p_1,p_2,\ldots p_n \in \mathbb{R}$

$$(p_1+p_2+\cdots+p_n)^m = \sum_{x\in A}\frac{m}{x_1!\,x_2!\,x_3!\cdots x_n!}p_1^{x_1}p_2^{x_2}\cdots p_n^{x_n}$$

$$(1)^m =$$

$$1 =$$

Consider the scenario where we choose one x_i as a success and all others as failures. Now x_i is the count of successes of m independent trials. Thus we'd expect $X_i \sim bin(m,p_i)$.

Lets check using X_n as a success.

$$f_{x_n}(x_n)=\sum_{x_1}\sum_{x_2}\sum_{x_3}\cdots\sum_{x_n}\frac{m!}{x_1!\,x_2!\cdots x_n!}p_1^{x_1}\cdots p_n^{x_n}$$

$$=\sum_{(x_1,x_2,x_3\ldots x_{n-1})\in B}\frac{m!}{x_1!\,x_2!\cdots x_n!}p_1^{x_1}\cdots p_n^{x_n}\cdot\left(\frac{(m-x_n)!\,(1-p_n)^{m-x_n}}{(m-x_n)!\,(1-p_n)^{m-x_n}}\right) \leftarrow \text{clever factor 1}$$

$$=\frac{m!}{x_n!\,(m-x_n)!}p_n^{x_n}(1-p_n)^{m-x_n}\times\underbrace{\sum_{(x_1,x_2,x_3\ldots x_{n-1})\in B}\frac{(m-x_n)!}{x_1!\,x_2!\cdots x_{n-1}!}\left(\frac{p_1}{1-p_n}\right)^{x_1}\cdots\left(\frac{p_{n-1}}{1-p_n}\right)^{x_{n-1}}}_{\text{Kernal}}$$

$$=\frac{m!}{x_n!\,(m-x_n)!}p_n^{x_n}(1-p_n)^{m-x_n}\sim Bin(m,p_n)$$

- Now let's consider the dist of failures

$$f_{(x_1, x_2, \ldots x_{n-1} | x_n)} = \frac{f(x_1, x_2 \ldots x_n)}{f(x_n)} = \text{something really ugly}$$

$$= \frac{m-x_n!}{x_1! x_2! \cdots x_{n-1}!} \left(\frac{p_1}{1-p_n}\right)^{x} \cdots \left(\frac{p_1}{1-p_n}\right)^{x_{n-1}}$$

↑ multinomial

- $Cov(X_i, X_j) = E[(X_i - p_i)(X_j - p_j)] = -mp_ip_j$

<u>Definition 4.6.5</u> Let $X_1, \ldots X_n$ be random vectors w/ joint pdf or pmf $f(x_1, x_2, \ldots x_n)$. Let $f_{X_i}(x_i)$ denote the marginal pdf or pmf of X_i then $X_1, \ldots X_n$ are called <u>mutually independent random vectors</u> if $\forall (x_1, x_2 \ldots x_n)$

$$f(x_1, x_2, \ldots x_n) = f_{x_1}(x_1) \cdot f_{x_2}(x_2) \cdot f_{x_3}(x_3) \cdots f_{x_n}(x_n) = \prod_{i}^{n} f_{x_i}(x_i)$$

* if $x_1, x_2, x_3, \ldots x_n$ are of dimension one we say mutually independent RV

<u>Theorem 4.6.6</u> Let $X_1, \ldots X_n$ be mutually independent RV. Let $g_1, \ldots g_n$ be real-valued functions $\ni$ $g_i(x_i)$ is a function only of x_i then

$$E(g_1(x_1) g_2(x_2) g_3(x_3) \cdots g_n(x_n)) = E(g_1(x_1)) E(g_2(x_2)) E(g_3(x_3)) \cdots E(g_n(x_n))$$

<u>Theorem 4.6.7</u> Let $X_1, \ldots X_n$ be mutually independent RV w/ mgfs $M_{x_1}(t), M_{x_2}(t), \ldots M_{x_n}(t)$. Let $Z = X_1 + \ldots + X_n$ then the Mgf is $M_Z(t) = M_{x_1}(t) M_{x_2}(t) \ldots M_{x_n}(t)$

Corollary 4.6.9 Let $X_1, \ldots, X_n$ be mutually independent RV w Mgfs $M_{X_1}(t), \ldots, M_{X_2}(t)$. Let $a_1, \ldots a_n$ and $b_1, \ldots b_n$ be fixed constants. Let $Z = (a_1X_1 + b_1) + (a_2X_2 + b_2) \cdots (a_nX_n + b_n)$

$M_Z(t) = e^{t(\Sigma b)} M_{X_1}(a_1 t)\, M_{X_2}(a_2 t) \cdots M_{X_n}(a_n t)$

Proof $M_Z(t) = E(e^{tZ}) = E\left(e^{t(\Sigma(a_i x_i + b_i))}\right) = e^{t(\Sigma b_i)} E\left(e^{t a_1 x_1} e^{t a_2 x_2} \cdots e^{t a_n x_n}\right)$

$= e^{t(\Sigma b)} M_{X_1}(t) \cdots M_{X_n}(t)$

Corollary 4.6.10 Let $X_1, X_2 \ldots X_n$ be mutually independent RV with $X_i \sim N(\mu_i, \sigma_i^2)$. Let $a_1, \ldots a_n$ and $b_1, \ldots b_n$ be fixed constants then

$$Z = \sum_{i=1}^{n} (a_i X_i + b_i) \sim N\left(\sum_{i=1}^{n} (a_i \mu_i + b_i), \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

Prove by product of MGFs

Theorem 4.6.11 Let $X_1, \ldots X_n$ be random vectors then $X_1, \ldots X_n$. They are mutually independent iff $\exists\, g_i(x)$ for $i = 1, \ldots n$ $\ni$ we can rewrite as

$$f(x_1, x_2, x_3, \ldots) = g_1(x_1) \cdot g_2(x_2) \cdots g_n(x_n)$$

Theorem 4.6.12 Let $X_1, \ldots, X_n$ be independent random vectors. Let $g_i(x_i)$ be a function only of x_i $i = 1, \ldots n$ random variables $U_i = g_i(X_i)$ $i = 1, \ldots, n$ are mutually ind

Example: $(X_1, \dots, X_n)$ w/ pdf $f_X(x_1, \dots x_n) = 24e^{-x_1-x_2-x_3-x_4}$

Let $A = \{x \mid f_X(x) > 0\}$

- Consider a new random vector $(U_1, \dots U_n)$ defined by $U_1 = g_1(X_1, \dots X_n)$ $U_2 = g_2(X_1, \dots, X_n) \dots U_n = g_n(X_1, \dots X_n)$. Suppose that $A_0, A_1, \dots A_k$ form a partition of A with these properties. The set A_0 which may be empty satisfies $P((X_1, X_2, \dots X_n) \in A_0) = 0$
- The transformation $(U_1, \dots U_n) = (g_1(x), \dots g_n(x))$ is a one to one transformation from A_i onto B $\forall$ $i = 1, 2 \dots k$ then $\forall$ i, the inverse functions from B to A_i can be found. Denote the ith inverse by $x_1 = h_{1i}(u_1, \dots, u_n)$, $x_2 = h_{2i}(u_1, \dots u_n), \dots x_n = h_{ni}(u_1, \dots, u_n)$. This ith inverse gives for $(u_1, \dots u_n) \in B$ the unique $(x_1, \dots x_n) \in A_i \ni$ $(u_1, \dots u_n) = (g_1(x_1, x_2, \dots x_n), \dots g_n(x_1, \dots x_n))$

$$J_i = \begin{vmatrix} \frac{dx_1}{du_1} & \frac{dx_1}{du_2} & \cdots & \frac{dx_1}{du_n} \\ \frac{dx_2}{du_1} & \frac{dx_2}{du_2} & \cdots & \frac{dx_2}{du_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dx_n}{du_1} & \frac{dx_n}{du_2} & \cdots & \frac{dx_n}{du_n} \end{vmatrix} \quad \text{where } x_{ji} = h_{ji}(u)$$

$$f_U(u_1, \dots u_n) = \sum_i^k f_X(h_{1i}(u_1, u_2, \dots u_n), h_{2i}(u_1, u_2, \dots u_n), \dots h_{ni}(u_1, u_2, \dots u_n)) |J_i|$$

Example

- Let $(X_1, X_2, X_3, X_4) \sim f_X(x_1, x_2, x_3, x_4) = 24e^{-x_1-x_2-x_3-x_4}$ $0 < x_1 < x_2 < x_3 < x_4 < \infty$
- Consider $U_1 = X_1$ $U_2 = X_2 - X_1$ $U_3 = X_3 - X_2$ $U_4 = X_4 - X_3$ all $0 < U_i < \infty$
- The transformation is one to one so $k = 1$

invs $\Rightarrow$ $X_1 = U_1$ $X_2 = U_1 + U_2$ $X_3 = U_1 + U_2 + U_3$ $X_4 = U_1 + U_2 + U_3 + U_4$

$$J = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{vmatrix} = 1$$

$$\begin{aligned} f_U(u_1, u_2, u_3, u_4) &= |J| f_X(u_1, u_1 + u_2, u_1 + u_2 + u_3, u_1 + u_2 + u_3 + u_4) \\ &= 24e^{-u_1 - u_1 - u_2 - u_1 - u_2 - u_3 - u_1 - u_2 - u_3 - u_4} \\ &= 24e^{-4u_1 - 3u_2 - 2u_3 - u_4} \end{aligned}$$

$\boxed{4.7}$ Lemma 4.7.1

Let a and b be any positive numbers, and let p and q be any positive numbers >1 if $\frac{1}{p}+\frac{1}{q}=1$ then $\frac{1}{p}a^p+\frac{1}{q}b^q \geq ab$ (equality when $a^p=b^q$)

Proof

- $g(a) = \frac{1}{p}a^p + \frac{1}{q}b^q - ab$ for fixed b
- $d/da = a^{p-1} - b$
- set $a^{p-1} - b = 0 \rightarrow a^{p-1} = b$
- plug b in: $\frac{1}{p}a^p + \frac{1}{q}(a^{p-1})^q - a^1 a^{p-1} = \frac{1}{p}a^p + \frac{1}{q}a^p - a^p$

$$= a^p\left(\frac{1}{p}+\frac{1}{q}\right) - a^p$$
$$= a^p - a^p$$
$$= 0$$

Theorem 4.7.2. Hölder's Inequality

Let X and Y be any two random variables, and let p and q satisfy (4.7.1) Then $|E(XY)| \leq E(|XY|) \leq (E(|X|)^p)^{1/p} E(|Y|^q)^{1/q}$

Proof $\quad -|XY| \leq XY \leq |XY|$

$a = \dfrac{|X|}{E(|X|^p)^{1/p}} \qquad b = \dfrac{|Y|}{E(|Y|^q)^{1/q}} \qquad$ Note: $a, b > 1$, $\frac{1}{p}+\frac{1}{q}=1$

$$\frac{1}{p}\left(\frac{|X|^p}{E(|X|^p)}\right) + \frac{1}{q}\left(\frac{|Y|^q}{q\,E(|Y|^q)}\right) \geq \frac{|XY|^p}{pq\,E(|X|^p)E(|Y|^q)}$$ by Lemma 4.7.1

Theorem 4.7.3 - Cauchy-Schwarz Inequality

For any two RV X and Y

$$|E(XY)| \leq E(|XY|) \leq (E(|X|^2))^{1/2}(E(|Y|^2))^{1/2}$$

Example 4.7.4 X and Y have means μ_x and μ_y and variances σ_x^2 and σ_y^2 we can apply 4.73

$$E(|(X-\mu_x)(Y-\mu_y)|) \le (E((X-\mu_x)^2))^{1/2}(E((Y-\mu_y)^2))^{1/2}$$

* square both sides

$$(E(|(X-\mu_x)(Y-\mu_y)|))^2 \le E((X-\mu_x)^2)E((Y-\mu_y)^2)$$

$$(Cov(X,Y))^2 \le \sigma_x^2\sigma_y^2$$

$$\left|\frac{cov(X,Y)}{\sigma_x\sigma_y}\right| \le 1$$

Aside Liapounov's Inequality $(E(|X|^r))^{1/r} \le (E(|X|^s)^{1/s})$ $1 < r < s < \infty$

Theorem 4.75 Minkowski's Inequality $\forall$ X and Y be 2 RVs $\ni$ $1 \le p \le \infty$

$$(E(|X+Y|^p)^{1/p}) \le (E(X)^p)^{1/p} + (E(|Y|^p))^{1/p}$$

Proof $E(|X+Y|^p) = E(|X+Y||X+Y|^{p-1})$

$\le E(|X||X+Y|^{p-1}) + E(|Y||X+Y|^{p-1})$

↑ by the triangle ineq. $|X+Y| \le |X|+|Y|$

$\le [(E(|X|^p))^{1/p}(E(|X+Y|^{q(p-1)})^{1/q}] + [(E(|Y|^p))^{1/p}(E(|X+Y|^{q(p-1)}))^{1/q}]$

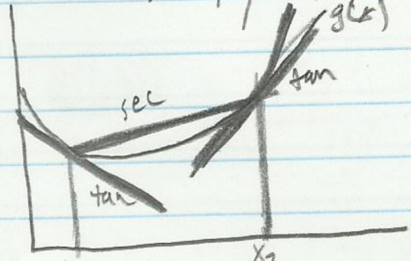
↑ by holders inequality

$E(|X+Y|^p)^{1/p} \le (E(|X|^p)^{1/p}) + (E(|Y|^p))^{1/p}$

↑ by dividing by $(E(|X+Y|^{q(p-1)}))^{1/q}$

Aside
- $\sum_{i=1}^{n}|a_ib_i| \le (\sum_{i=1}^{n} a_i^p)^{1/p}(\sum_{i=1}^{n} b_i^q)^{1/q}$ $\frac{1}{p}+\frac{1}{q}=1$
- Special case $\frac{1}{n}(\sum_{i=1}^{n}|a_i|)^2 \le \sum_{1}^{n} a^2$

Definition A function $g(x)$ is convex if $g(\lambda x+(1-\lambda)y) \le \lambda g(x)+(1-\lambda)g(y)$ $\forall$ x, y and $0 < \lambda < 1$. The function $g(x)$ is concave if $-g(x)$ is convex.



Convex function w/ tangent lines at x_1, x_2 and secant line

Theorem 4.7.7 __Jensen's Inequality__ for any RV X if $g(x)$ is a convex function then $E(g(x)) \geq g(E(x))$ (equality holds iff $\forall$ line $a+bx$ tangent to $g(x)$ at $x = E(x)$ $P(g(x) = a+bx) = 1$

- Shows $E(X^2) \geq (E(x))^2$ as x^2 is convex
- Shows $E(1/x) \geq \frac{1}{E(x)}$ for $x > 0$ as $1/x$ is convex

__Second derivative test__ For twice differentiable $g(x)$

$$g''(x) \geq 0 \rightarrow \text{convex} \qquad g''(x) \leq 0 \rightarrow \text{convex}$$

__Corollary__ In theorem 4.7.7 we can add if g is concave $E(g(x)) \leq g(E(x))$

__Example:__ Let $a_1, a_2, \ldots a_n \geq 0$

$a_A = \frac{1}{n} \sum_{i=1}^{n} a_i$ (arithmetic mean)

$a_G = \left[\prod_{i}^{n} a_i\right]^{1/n}$ (geometric mean)

$a_H = \frac{1}{\frac{1}{n}\left[\prod \frac{1}{a_i}\right]}$ (harmonic mean)

$$a_H \leq a_G \leq a_A$$

- Let X be a RV w/ range $a_1, \ldots, a_n$ and $P(X = a_i) = \frac{1}{n}$
- We can consider a concave function: $\log(x)$
- Jensen's inequality shows $E(\log(X)) \leq \log(E(x))$
- So $\log(a_G) = \frac{1}{n} \sum_{i=1}^{n} \log(a_i) = E(\log X) \leq \log(E(x)) = \log(\frac{1}{n} \sum_1 a_i) = \log a_A$ so $a_G \leq a_A$
  - similar logic for other parts p191

- If X is a RV w/ finite mean μ and $g(x)$ is non decreasing $E(g(x)(x - \mu)) \geq 0$

Theorem 4.7.9 Let X be any RV and g(x) and h(x) any functions $\ni$ $E(g(X))$, $E(h(X))$ and $E(g(X)h(X))$ exist

a) If g(x) is a nondecreasing function and h(x) is non decreasing then
$E(g(X)h(X)) \leq E(g(X))E(h(X))$

b) If g(x), h(x) are both non dec or non inc
$E(g(X)(h(X)) \geq E(g(X))E(h(X))$