

richardwu.ca

# STAT 330 COURSE NOTES

MATHEMATICAL STATISTICS

YINGLI QIN • FALL 2018 • UNIVERSITY OF WATERLOO

---

Last Revision: September 19, 2018

## Table of Contents

<b>1</b>	<b>September 7, 2018</b>	<b>1</b>
1.1	Random variables . . . . .	1
<b>2</b>	<b>September 10, 2018</b>	<b>3</b>
2.1	Cumulative distribution function (cdf) . . . . .	3
2.2	Location parameters . . . . .	3
2.3	Scale parameters . . . . .	4
2.4	Pivotal quantities . . . . .	4
<b>3</b>	<b>September 12, 2018</b>	<b>5</b>
3.1	Pdf of a function . . . . .	5
<b>4</b>	<b>September 14, 2018</b>	<b>7</b>
4.1	Expectations . . . . .	7
4.2	Markov's inequality . . . . .	7
4.3	Moment generating function (mgf) . . . . .	8
<b>5</b>	<b>September 17, 2018</b>	<b>9</b>
5.1	Derivatives of mgf . . . . .	9
5.2	Joint cdf and pdf . . . . .	10

---

### Abstract

These notes are intended as a resource for myself; past, present, or future students of this course, and anyone interested in the material. The goal is to provide an end-to-end resource that covers all material discussed in the course displayed in an organized manner. These notes are my interpretation and transcription of the content covered in lectures. The instructor has not verified or confirmed the accuracy of these notes, and any discrepancies, misunderstandings, typos, etc. as these notes relate to course's content is not the responsibility of the instructor. If you spot any errors or would like to contribute, please contact me directly.

## 1 September 7, 2018

### 1.1 Random variables

We have two types (not include mixture r.v.s) random variables (r.v.s):

**Discrete** Probability (mass) function of  $X$

$$f(x) = P(X = x)$$

Support set of  $X$

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

The following two conditions must hold

•

$$f(x) \geq 0$$

•

$$\sum_{x \in A} f(x) = 1 \quad \text{or} \quad \sum_{x \in \mathbb{R}} f(x) = 1$$

**Continuous** Probability density function (pdf) of  $X$

$$f(x) = \frac{d}{dx} F(x) = F'(x)$$

if  $F$  is differentiable at  $x$ , otherwise  $f(x) = 0$ .

Support set of  $X$

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

The following two conditions must hold

•

$$f(x) \geq 0 \quad \forall x \in \mathbb{R}$$

•

$$\int_{x \in A} f(x) dx = 1 \quad \text{or} \quad \int_{-\infty}^{\infty} f(x) dx = 1$$

Some examples of **discrete** r.v.s

**Bernoulli**  $X \sim \text{Bernoulli}(p)$  for  $0 < p < 1$  where

$$P[X = 1] = p \quad \text{or} \quad P[X = 0] = 1 - p$$

therefore

$$f(x) = P[X = x] = p^x(1-p)^{1-x} \quad x = 0, 1$$

and  $A = \{0, 1\}$ .

**Binomial**  $X \sim \text{BIN}(n, p)$  for  $n = 1, 2, \dots$  and  $0 < p < 1$ .  $X$  represents the number of successes of  $n$  iid  $\text{BERN}(p)$  trials or  $X$  (or  $X$  is sum of  $n$  iid  $\text{BERN}(p)$  r.v.s):

$$X = \sum_{i=1}^n Y_i \quad Y_i \sim \text{BERN}(p)$$

therefore

$$f(x) = P[X = x] = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \dots, n$$

and  $A = \{1, 2, \dots, n\}$ .

**Geometric**  $X \sim \text{GEO}(p)$  for  $0 < p < 1$ .  $X$  represents the number of failures before the 1st success in a sequence of iid  $\text{BERN}(p)$  trials, therefore

$$f(x) = P[X = x] = (1-p)^x p \quad x = 0, 1, \dots$$

and  $A = \{0, 1, \dots\}$ .

**Negative Binomial**  $X \sim \text{NB}(k, p)$  where  $X$  represents the number of successes in  $k$   $\text{BERN}(p)$  trials. We skip this for now.

Some examples of **continuous** r.v.s

**Normal/Gaussian**  $X \sim N(\mu, \sigma^2)$  for  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ .

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in \mathbb{R}$$

**Gamma**  $X \sim \text{GAM}(\alpha, \beta)$  for  $\alpha, \beta > 0$ . The pdf may be left or right skewed.

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) \quad x \in \mathbb{R}^+$$

Note that the Gamma function  $\Gamma$  is defined as

$$\begin{aligned} \Gamma(\alpha) &= (\alpha-1)\Gamma(\alpha-1) \quad \alpha > 1 \\ \Gamma(n) &= (n-1)! \\ \Gamma\left(\frac{1}{2}\right) &= \sqrt{\pi} \end{aligned}$$

**Exponential**  $X \sim \text{EXP}(\theta)$  for  $\theta > 0$ .

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \quad x \geq 0$$

Note that  $\text{EXP}(\theta)$  is simply  $\text{GAM}(1, \theta)$ .

## 2 September 10, 2018

### 2.1 Cumulative distribution function (cdf)

We denote the *cumulative distribution function* (cdf) as  $F(x) = P[X \leq x]$  with properties:

1. non-decreasing i.e.  $F(a) \leq F(b)$  if  $a \leq b$

2.

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

3.

$$\lim_{x \rightarrow \infty} F(x) = 1$$

4. right-continuous, i.e.  $\lim_{x \downarrow x_0} F(x) = F(x_0)$  (where  $x \downarrow x_0$  denotes  $x$  approaches  $x_0$  from  $x_0$ 's right-hand side or in this case from above).

**Remark 2.1.** If  $X$  is a continuous r.v then  $F(x)$  is also left-continuous i.e.  $F(x)$  is continuous.

### 2.2 Location parameters

**Example 2.1.** If  $X \sim N(\mu, 1)$ ,  $\mu \in \mathbb{R}$ , then  $\mu$  is a location parameter for  $X$  where

$$f(x; \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} \quad x \in \mathbb{R}$$

$f(x, \mu)$  is *NOT completely specified* as  $f(\cdot, \mu)$  cannot be calculated at  $x$  as  $\mu$  is *unknown* (we would need to perform *statistical inference* to estimate  $\mu$ ).

On the other hand,  $f(x; 0)$  is completely specified. Notice that

$$\begin{aligned} f(x; \mu) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu-0)^2}{2}} \\ &= f(x - \mu; 0) \end{aligned}$$

That is: the uncompletely specified  $f(x; \mu)$  can be rewritten as a completely specified  $f(\cdot; 0)$  evaluated at  $x - \mu$ .  $\mu$  is a *location parameter* for  $X \sim N(\mu, 1)$ .

**Definition 2.1.** A quantity  $\eta$  is a **location parameter** for  $X$  with a pdf  $f(x; \eta)$  if

$$f(x; \eta) = f(x - \eta; 0)$$

Increasing the value of the location parameter of the pdf shifts it to the right (e.g. for  $N(\mu, 1)$ ).

For a continuous r.v.  $X$  with a location parameter  $\eta$

$$\begin{aligned} F(x; \eta) &= P[X \leq x; \eta] \\ &= \int_{-\infty}^x f(t; \eta) dt \\ &= \int_{-\infty}^x f(t - \eta; 0) dt \end{aligned}$$

since  $\eta$  is a location parameter for our pdf  $f$ . Let  $s = t - \eta$ , then

$$\begin{aligned} &= \int_{-\infty}^{x-\eta} f(s; 0) \, ds \\ &= F(x - \eta; 0) \end{aligned}$$

Therefore  $\eta$  is a location parameter iff  $F(x; \eta) = F(x - \eta; 0)$ .

### 2.3 Scale parameters

**Example 2.2.** Let  $X \sim EXP(\theta)$ ,  $\theta > 0$  (as we will see,  $\theta$  is a scale parameter for  $X$ ). Recall

$$f(x; \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \quad \theta > 0$$

is *NOT completely specified* as  $\theta$  is unknown.

However  $f(x; 1) = \exp(-x)$  for  $x > 0$  is the pdf of  $EXP(1)$  which is completely satisfied. Note that

$$f(x; \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right) = \frac{1}{\theta} f\left(\frac{x}{\theta}; 1\right)$$

$\theta$  is a *scale parameter* for  $X \sim EXP(\theta)$ ,  $\theta > 0$ .

**Definition 2.2.** A quantity  $\theta$  is a **scale parameter** if its pdf satisfies

$$f(x; \theta) = \frac{1}{\theta} f\left(\frac{x}{\theta}; 1\right) \quad \theta > 0$$

That is: the uncompletely specified pdf can be re-written as the product of  $\frac{1}{\theta}$  and a completely specified pdf  $f(\cdot; 1)$  evaluated at  $\frac{x}{\theta}$ .

How about the corresponding cdf (for a continuous r.v with scale parameter  $\theta$ )?

$$\begin{aligned} F(x; \theta) &= \int_{-\infty}^x f(t; \theta) \, dt \\ &= \int_{-\infty}^x f\left(\frac{t}{\theta}; 1\right) \frac{1}{\theta} \, dt \end{aligned}$$

since  $\theta$  is a scale parameter. Let  $s = \frac{t}{\theta}$  (so  $ds = \frac{dt}{\theta}$ ), thus

$$\begin{aligned} &= \int_{-\infty}^{\frac{x}{\theta}} f(s; 1) \, ds \\ &= F\left(\frac{x}{\theta}; 1\right) \end{aligned}$$

Therefore  $\theta$  is a scale parameter iff  $F(x; \theta) = F\left(\frac{x}{\theta}; 1\right)$ .

### 2.4 Pivotal quantities

**Remark 2.2.** If  $\eta$  is a location parameter, then  $\hat{\eta} - \eta$  is a pivotal quantity for constructing a confidence interval for  $\eta$  (where  $\hat{\eta}$  is the Maximum Likelihood Estimate (MLE) of  $\eta$ ).

If  $\theta$  is a scale parameter, then  $\frac{\hat{\theta}}{\theta}$  is a pivotal quantity for construct a confidence interval for  $\theta$ .

### 3 September 12, 2018

#### 3.1 Pdf of a function

We want to find the pdf of a function of one r.v.

**Method 1** Let  $Y = h(X)$ . If  $h(\cdot)$  is a **1-1 function** then  $h(\cdot)$  is either strictly increasing or strictly decreasing.

1. When  $h(\cdot)$  is strictly increasing ( $h^{-1}(\cdot)$  exists and is also strictly increasing): let  $G(y)$  be the cdf of  $Y$  and  $g(y)$  be the pdf of  $Y$ .

Given that  $X$  is a continuous r.v. with pdf  $f(x)$  and cdf  $F(x)$ , then

$$G(y) = P[Y \leq y] = P[h(X) \leq y] = P[X \leq h^{-1}(y)] = F(h^{-1}(y))$$

For the pdf  $g(y)$ , we have

$$\begin{aligned} g(y) &= \frac{dG(y)}{dy} = \frac{dF(h^{-1}(y))}{dy} \\ &= f(h^{-1}(y)) \cdot \frac{\partial h^{-1}(y)}{\partial y} \\ &= f(h^{-1}(y)) \cdot \left| \frac{\partial h^{-1}(y)}{\partial y} \right| \end{aligned}$$

since  $h^{-1}(\cdot)$  is strictly increasing, we have  $\frac{\partial h^{-1}(y)}{\partial y} > 0$  (so we can add an absolute sign).

2. When  $h(\cdot)$  and thus  $h^{-1}(\cdot)$  is strictly decreasing we have

$$\begin{aligned} G(y) &= P[h(X) \leq y] = P[h^{-1}(h(X)) \geq h^{-1}(y)] \\ &= P[X \geq h^{-1}(y)] \\ &= 1 - P[X < h^{-1}(y)] \\ &= 1 - P[X \leq h^{-1}(y)] & P[X = h^{-1}(y)] = 0 \text{ since } X \text{ is continuous} \\ &= 1 - F(h^{-1}(y)) \end{aligned}$$

For the pdf  $g(y)$

$$\begin{aligned} g(y) &= \frac{dG(y)}{dy} = \frac{d(1 - F(h^{-1}(y)))}{dy} \\ &= -f(h^{-1}(y)) \cdot \frac{\partial h^{-1}(y)}{\partial y} \\ &= f(h^{-1}(y)) \cdot \left| \frac{\partial h^{-1}(y)}{\partial y} \right| \end{aligned}$$

since  $h^{-1}(\cdot)$  is strictly decreasing thus  $\frac{\partial h^{-1}(y)}{\partial y} < 0$ , hence the absolute sign.

So if  $h(\cdot)$  is a **1-1 function**, we have for  $Y = h(X)$  the pdf

$$g(y) = f(h^{-1}(y)) \cdot \left| \frac{\partial h^{-1}(y)}{\partial y} \right|$$

How do we find the support set for  $Y$ ? Let  $A$  be the support set of  $X$  and  $B$  be the support set for  $Y$ . Let  $h : A \rightarrow B^*$  where  $B^*$  is the image of  $A$  under  $h(\cdot)$ .

Thus we have  $B = \{y \mid y \in B^* \text{ and } g(y) > 0\}$ .

**Example 3.1.** Let  $X$  have a pdf  $f(x) = \frac{\theta}{x^{\theta+1}}$  where  $x \geq 1$  and  $\theta > 0$ .

Find the pdf of  $Y = \log X$  (natural log).

We have  $h(X) = \log X$  thus  $X = e^Y = h^{-1}(Y)$ . Since  $h(x)$  is 1-1 we can use our previous result:

$$f(h^{-1}(y)) = f(e^y) = \frac{\theta}{(e^y)^{\theta+1}}$$

Also

$$\frac{\partial h^{-1}(y)}{\partial y} = \frac{\partial e^y}{\partial y} = e^y$$

Thus we have

$$\begin{aligned} g(y) &= \frac{\theta}{e^{y\theta} e^y} \cdot |e^y| \\ &= \frac{\theta}{e^{y\theta} e^y} \cdot e^y \\ &= \frac{\theta}{e^{y\theta}} \end{aligned}$$

To find the support, note that  $h(x) = \log X$  has support  $A = \{x \mid x \geq 1\}$  thus  $h : A \rightarrow B^* = \{y \mid y \geq 0\}$ . Note that  $g(y) = \frac{\theta}{e^{y\theta}} > 0$  for all  $y \in \mathbb{R}$ , thus the support for  $Y$  is  $B = B^* = \{y \mid y \geq 0\}$ .

**Method 2** For functions  $h(\cdot)$  that are not 1-1, we use the cdf technique.

**Example 3.2.** Let  $X \sim N(0, 1)$  and  $Y = X^2$ : find the pdf  $G(Y)$  of  $Y$ .

$$G(y) = P[Y \leq y] = P[X^2 \leq y]$$

Note that  $P[X^2 \leq 0] = P[X^2 = 0] = 0$  since  $x^2 \geq 0$  for all  $x \in \mathbb{R}$ , so if  $y = 0$  then  $G(y) = 0$ .

For  $y > 0$ , we have

$$\begin{aligned} G(y) &= P[X^2 \leq y] \\ &= P[-\sqrt{y} \leq X \leq \sqrt{y}] \\ &= 2P[0 \leq X \leq \sqrt{y}] && N(0, 1) \text{ is symmetric} \\ &= 2 \int_0^{\sqrt{y}} f(x) dx \\ &= 2 \int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \end{aligned}$$

We require  $g(y) = \frac{dG(y)}{dy}$ .

From Fundamental Theorem of Calculus, if  $f(x)$  is cont. on  $[a, b]$  and  $g(x) = \int_a^x f(t) dt \forall x \in [a, b]$  is cont. on  $[a, b]$  then

$$\frac{dg(x)}{dx} = f(x) \quad \forall x \in [a, b]$$

Thus for all  $y > 0$  we have

$$\begin{aligned}\frac{dG(y)}{dy} &= \frac{2}{\sqrt{2\pi}} \frac{d \int_0^{\sqrt{y}} e^{-\frac{x^2}{2}} dx}{dy} \\ &= \frac{2}{\sqrt{2\pi}} e^{-\frac{(\sqrt{y})^2}{2}} \cdot \frac{d\sqrt{y}}{dy} \\ &= -\frac{1}{\sqrt{\pi y}} e^{-\frac{y}{2}}\end{aligned}$$

So  $g(y) = \frac{1}{\sqrt{\pi y}} e^{-\frac{y}{2}}$  is the pdf of  $Y \sim X^2(1)$

Note that  $h : A \rightarrow B^*$  where  $A = \mathbb{R}$ , thus  $B^* = \{y \mid y > 0\}$ .

The support set of  $Y$  is  $B$  where  $B = \{y \mid y \in B^* \text{ and } g(y) > 0\}$ .

Notice that  $G(y) = 0$  if  $y = 0$  and  $G(y)$  is not differentiable at  $y = 0$ , thus  $g(0) = 0$  so  $B = \{y \mid y > 0\}$ .

## 4 September 14, 2018

### 4.1 Expectations

The expectation  $E(X)$  of a r.v.  $X$  exists if  $E(|X|) < \infty$ . It is defined as

**Discrete r.v.  $X$**

$$E(X) = \sum_{x \in A} x \cdot f(x)$$

By the Law of the Unconscious Statistician (LOTUS)

$$E(h(X)) = \sum_{x \in A} h(x) \cdot f(x)$$

**Continuous r.v.  $X$**

$$\begin{aligned}E(X) &= \int_A x f(x) dx \\ &= \int_{-\infty}^{\infty} x f(x) dx\end{aligned}$$

LOTUS holds for continuous r.v.'s as well

$$E(h(X)) = \int_A h(x) \cdot f(x) dx$$

### 4.2 Markov's inequality

**Theorem 4.1** (Markov's inequality). Markov's inequality states that

$$P[|X| \geq c] \leq \frac{E[|X|^k]}{c^k}$$

for all  $c, k > 0$ .



*Proof.* Note that  $P[|X| \geq c] = P[X \leq -c] + P[X \geq c]$  or the tail probabilities beyond  $-c$  and  $c$ .

Thus Markov's inequality gives an *upper bound* for the tail probabilities.

In the continuous case we have for the RHS

$$P[|X| \geq c] = \int_{\{|x||x| \geq c\}} f(x) dx$$

For the LHS we have

$$\begin{aligned} \frac{E[|X|^k]}{c^k} &= E\left[\left|\frac{X}{c}\right|^k\right] = \int_{-\infty}^{\infty} \left|\frac{x}{c}\right|^k f(x) dx \\ &= \int_{|x| \geq c} \left|\frac{x}{c}\right|^k f(x) dx + \int_{|x| < c} \left|\frac{x}{c}\right|^k f(x) dx \\ &\geq \int_{|x| \geq c} \left|\frac{x}{c}\right|^k f(x) dx && \text{right term is integral over non-negative function} \\ &\geq \int_{|x| \geq c} f(x) dx && |x| \geq c \Rightarrow \left|\frac{x}{c}\right|^k \geq 1 \end{aligned}$$

and the result follows.  $\square$

**Example 4.1.** Given  $X \sim N(0, \sigma^2)$ , what is a bound on  $P[|X| \geq 3\sigma]$ ?

From Markov's inequality, let  $k = 2$  (where  $E[X^2] = \sigma^2$ )

$$\begin{aligned} P[|X| \geq 3\sigma] &\leq \frac{E[|X|^k]}{(3\sigma)^k} \\ &= \frac{E[X^2]}{9\sigma^2} \\ &= \frac{\sigma^2}{9\sigma^2} \\ &= \frac{1}{9} \end{aligned}$$

Since  $P[|X| \geq 3\sigma] \leq \frac{1}{9}$  then  $P[|X| \leq 3\sigma] \geq 1 - \frac{1}{9} = \frac{8}{9}$ .

Thus  $X$  stays  $3\sigma$  distance from 0 with a high probability of at least  $\frac{8}{9}$ .

### 4.3 Moment generating function (mgf)

**Definition 4.1** (Moment generating function). For a r.v.  $X$  the expectation

$$M_X(t) = E[e^{tX}]$$

is called the moment generating function (if the expectation exists).

One must state the values of  $t$  such that  $M_X(t)$  exists ("domain of convergence").

**Example 4.2.** Let  $X \sim GAM(\alpha, \beta)$ ,  $\alpha, \beta > 0$ . Find  $M_X(t)$ .

$$\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= \int_0^\infty e^{tx} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} dx \\
&= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-x(\frac{1}{\beta}-t)} dx
\end{aligned}$$

Note that for any pdf  $f(x)$  we have  $\int_A f(x) dx = 1$ , thus  $\int_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} = 1$  thus  $\int_0^\infty x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \beta^\alpha \Gamma(\alpha)$ . Thus we have from before

$$\begin{aligned}
\frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-x(\frac{1}{\beta}-t)} dx &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \left( \frac{1}{\frac{1}{\beta}-t} \right)^\alpha \Gamma(\alpha) \\
&= \frac{1}{(1-\beta t)^\alpha}
\end{aligned}$$

when  $(\frac{1}{\beta}-t)^{-1} > 0$  i.e.  $t < \frac{1}{\beta}$ . What if  $t \geq \frac{1}{\beta}$ ? When  $t = \frac{1}{\beta}$  our integral becomes  $\int_0^\infty x^{\alpha-1} dx$  which goes to infinity for  $\alpha > 0$ .

Similarly it goes to infinity when  $t > \frac{1}{\beta}$ .

## 5 September 17, 2018

### 5.1 Derivatives of mgf

For the continuous case (similarly for discrete) we can take the derivative of the mgf  $M_X(t)$

$$\begin{aligned}
\frac{dM_X(t)}{dt} &= \frac{d}{dt} \sum_{-\infty}^{\infty} e^{tX} f(x) dx \\
&= \sum_{-\infty}^{\infty} \frac{d}{dt} [e^{tX} f(x)] dx && \text{Leibniz rule} \\
&= \sum_{-\infty}^{\infty} x e^{tX} f(x) dx
\end{aligned}$$

We can clearly see when  $t = 0$  we have the expected value  $E[X]$ . Similarly

$$\begin{aligned}
\frac{d^2 M_X(t)}{dt^2} &= \frac{d}{dt} \left[ \frac{d}{dt} M_X(t) \right] \\
&= \frac{d}{dt} \left[ \sum_{-\infty}^{\infty} x e^{tX} f(x) dx \right] \\
&= \sum_{-\infty}^{\infty} \frac{d}{dt} [x e^{tX} f(x)] dx \\
&= \sum_{-\infty}^{\infty} x^2 e^{tX} f(x) dx
\end{aligned}$$

which we recognize when  $t = 0$  as the second moment  $E[X^2]$ .

In summary

$$\frac{d^r}{dt^r} M_X(t) = \int_{-\infty}^{\infty} x^r e^{tX} f(x) dx \quad r = 1, 2, \dots$$

where

$$\begin{aligned} \left( \frac{d^r}{dt^r} M_X(t) \right) \Big|_{t=0} &= \left( \int_{-\infty}^{\infty} x^r e^{tX} f(x) dx \right) \Big|_{t=0} \\ &= \int_{-\infty}^{\infty} x^r f(x) dx \\ &= E[X^r] \end{aligned}$$

**Example 5.1.** For  $X \sim \text{GAM}(\alpha, \beta)$  we have  $M_X(t) = \frac{1}{(1-\beta t)^\alpha}$ ,  $t < \frac{1}{\beta}$ . Find  $E[X]$  and  $\text{Var}(X)$ . Note that  $\text{Var}(X) = E[X^2] - E[X]^2$ . Also

$$\begin{aligned} \frac{d}{dt} M_X(t) &= \frac{d}{dt} [(1 - \beta t)^{-\alpha}] \\ &= (-\alpha)(-\beta)(1 - \beta t)^{-\alpha-1} \\ &= \alpha\beta(1 - \beta t)^{-\alpha-1} \end{aligned}$$

Thus  $E[X] = \alpha\beta(1 - \beta 0)^{-\alpha-1} = \alpha\beta$ .

Similarly  $E[X^2] = \alpha(\alpha + 1)\beta^2$  thus  $\text{Var}(X) = \alpha\beta^2$ .

## 5.2 Joint cdf and pdf

The joint cdf  $F[x, y]$  is defined as  $P[X \leq x \text{ and } Y \leq y]$  or simply  $P[X \leq x, Y \leq y]$ .

Recall that for the cdf  $F(x)$  for  $X$ , we have

1.  $F(a) \leq F(b)$  if  $a \leq b$
2.  $\lim_{x \rightarrow -\infty} F(x) = 0$
3.  $\lim_{x \rightarrow \infty} F(x) = 1$
4.  $\lim_{x \downarrow x_0} F(x) = F(x_0)$  (right continuous)

Similarly, the properties for the *joint cdf* of  $X$  and  $Y$  are

1. For every fixed  $y$ ,  $F(x, y)$  is non-decreasing for  $x$ . Similarly for fixed  $x$ ,  $F(x, y)$  is non-decreasing for  $y$ .
2. For every fixed  $y$ ,  $\lim_{x \rightarrow -\infty} F(x, y) = 0$  (similarly with fixed  $x$  and  $y \rightarrow -\infty$ ).
3.  $\lim_{x, y \rightarrow -\infty} F(x, y) = 0$
4.  $\lim_{x, y \rightarrow \infty} F(x, y) = 1$

$$F_1(x) = P[X \leq x] = \lim_{y \rightarrow \infty} F(x, y)$$

$$F_2(y) = P[Y \leq y] = \lim_{x \rightarrow \infty} F(x, y)$$

Comparing discrete and continuous joint r.v.s

**Discrete r.v.** For the pmf we have

$$f(x, y) = P(X = x, Y = y)$$

Our support set is  $A = \{(x, y) \mid f(x, y) > 0\}$ .

For the pmf, we have

1.  $f(x, y) \geq 0$  for all  $(x, y) \in \mathbb{R} \times \mathbb{R}$
2.  $\sum \sum f(x, y) = 1$  where  $(x, y) \in A$

**Continuous r.v.** For the pdf we have

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

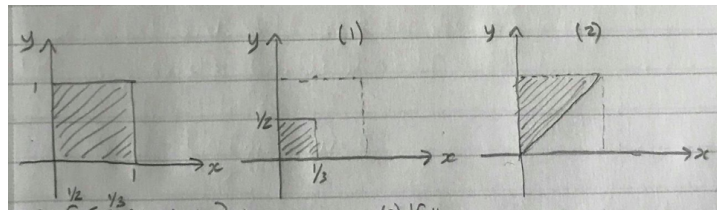
Our support set is  $A = \{(x, y) \mid f(x, y) > 0\}$ .

For the pdf, we have

1.  $f(x, y) \geq 0$  for all  $(x, y) \in \mathbb{R} \times \mathbb{R}$
2.  $\int \int f(x, y) dx dy = 1$  where  $(x, y) \in A$

**Example 5.2.** Suppose that  $X$  and  $Y$  are cont. r.v.s with joint pdf  $f(x, y) = x + y$  for  $0 < x < 1$  and  $0 < y < 1$ . Find

1.  $P[X \leq \frac{1}{3}, Y \leq \frac{1}{2}] = F(\frac{1}{3}, \frac{1}{2})$
2.  $P[X \leq Y]$
3.  $P[X + Y \leq \frac{1}{2}]$



**Figure 5.1:** Diagram of area we are trying to integrate over for (1) and (2).

**Solution.** 1. We sum over the shaded square area

$$\begin{aligned} \int_0^{\frac{1}{2}} \left( \int_0^{\frac{1}{3}} f(x, y) dx \right) dy &= \int_0^{\frac{1}{2}} \left( \int_0^{\frac{1}{3}} x + y dx \right) dy \\ &= \int_0^{\frac{1}{2}} \left( \frac{x^2}{2} + xy \Big|_{x=0}^{x=1/3} \right) dy \\ &= \int_0^{\frac{1}{2}} \frac{1}{18} + \frac{y}{3} dy \\ &= \frac{y}{18} + \frac{y^2}{6} \Big|_{y=0}^{y=1/2} \\ &= \frac{5}{72} \end{aligned}$$

2. If the region is not rectangular we pick one variable first, say  $y$ , and range from its smallest value to the largest value in its region.

We then find the range of the other variable ( $x$  in this case) for every given  $y$ .

$$\begin{aligned} P[X \leq Y] &= \int_0^1 \left( \int_0^y f(x, y) \, dx \right) dy \text{ OR} \\ &= \int_0^1 \left( \int_x^1 f(x, y) \, dy \right) dx \end{aligned}$$