

richardwu.ca

STAT 330 COURSE NOTES

MATHEMATICAL STATISTICS

YINGLI QIN • FALL 2018 • UNIVERSITY OF WATERLOO

Last Revision: October 17, 2018

Table of Contents

1	September 7, 2018	1
1.1	Random variables	1
2	September 10, 2018	3
2.1	Cumulative distribution function (cdf)	3
2.2	Location parameters	3
2.3	Scale parameters	4
2.4	Pivotal quantities	4
3	September 12, 2018	5
3.1	Pdf of a function	5
4	September 14, 2018	7
4.1	Expectations	7
4.2	Markov's inequality	7
4.3	Moment generating function (mgf)	8
5	September 17, 2018 and September 19, 2018	9
5.1	Derivatives of mgf	9
5.2	Joint cdf and pdf	10
6	September 21, 2018	14
6.1	Independence	14
6.2	Factorization independence theorem	15
7	September 24, 2018	16
7.1	Conditional pmf/pdf	16
7.2	Product rule	16
7.3	Independence and condition pmf/pdfs	18
8	September 26, 2018	19
8.1	Joint expectation	19
8.2	Conditional expectation	19
8.3	Expectation of a conditional expectation	21
8.4	Variance as sum of conditional expectations	21

9 September 28, 2018	22
9.1 Joint moment generating function (mgf)	22
9.2 Independence and joint mgfs	23
9.3 Summary of methods for verifying independence	23
10 October 1, 2018	24
10.1 Expectations/moments from mgf	24
10.2 Multinomial distribution	25
10.3 Mgf of multinomial distribution	26
10.4 Subset of multinomial is multinomial	26
11 October 3, 2018	27
11.1 More multinomial problems	27
11.2 Bivariate normal distribution	28
12 October 5, 2018	29
12.1 Remarks of bivariate normal	29
12.2 Mgf of bivariate normal	30
12.3 Joint cdf from pdf	31
13 October 12, 2018	31
13.1 Marginal pdf of bivariate normal	31
14 October 15, 2018	32
14.1 Bivariate transformation	32
15 October 17, 2018	37
15.1 Verifying 1-to-1 for bivariate functions	37
15.2 Examples of bivariate transformations	37
15.3 Computing determinant from the inverse's determinant	39
15.4 Bivariate transformations of non 1-to-1 functions	39

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 Γ 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 μ 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 μ is *unknown* (we would need to perform *statistical inference* to estimate μ).

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 $x - \mu$. μ is a *location parameter* for $X \sim N(\mu, 1)$.

Definition 2.1. A quantity η 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 η

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

since η 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 η 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, θ 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 θ is unknown.

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

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

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

Definition 2.2. A quantity θ 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 θ)?

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

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

2.4 Pivotal quantities

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

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

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. □

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 3σ 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 $\int_{-\infty}^\infty x^{\alpha-1} dx$ which goes to infinity for $\alpha > 0$.

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

5 September 17, 2018 and September 19, 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. $\lim_{x, y \rightarrow \infty} F(x, y) = 1$
- 4.

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

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

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 $(x, y) \in A$

To compute the marginal pmf for x we take

$$f_1(x) = \sum_{y \in \mathbb{R}} f(x, y)$$

(similarly for the marginal pmf for y).

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 $(x, y) \in A$

To compute the marginal pdf for x we take

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

(similarly for the marginal pdf for y).

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. $P[X + Y \leq \frac{1}{2}]$
4. $P[XY \leq \frac{1}{2}]$
5. $f_1(x)$
6. $F(x, y)$
7. $F_1(x)$

Solution. Note that while we may be finding $P[X \leq \frac{1}{3}]$ which is generally everything to the right of $x = \frac{1}{3}$, we only want the region intersected by our support set. This is represented as the shaded region in the diagrams below.

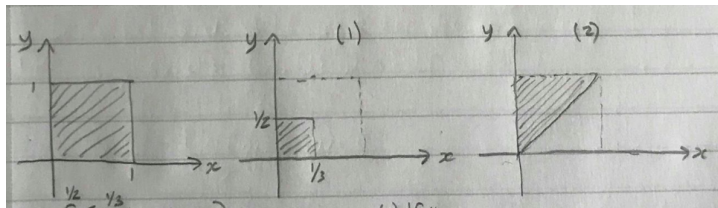


Figure 5.1: Diagram of area we are trying to integrate over for (1) and (2).

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 .

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

We have

$$\begin{aligned}
 P[X \leq Y] &= \int_0^1 \left(\int_0^y f(x, y) dx \right) dy \\
 &= \int_0^1 \left(\int_0^y x + y dx \right) dy \\
 &= \int_0^1 \frac{3y^2}{2} dy \\
 &= \frac{1}{2}
 \end{aligned}$$

3. The region is the triangle under the line $y = \frac{1}{2} - x$ in quadrant 1.

$$\begin{aligned}
 P[X + Y \leq \frac{1}{2}] &= P[Y \leq -x + \frac{1}{2}] \\
 &= \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}-y} x + y \, dx \, dy \\
 &= \int_0^{\frac{1}{2}} \frac{x^2}{2} + xy \Big|_{x=0}^{\frac{1}{2}-y} \, dy \\
 &\vdots \\
 &= \frac{1}{24}
 \end{aligned}$$

4. We have $1 - XY \geq \frac{1}{2}$ thus

$$\begin{aligned}
 P[XY \geq \frac{1}{2}] &= \int_{\frac{1}{2}}^1 \int_{\frac{1}{2y}}^1 f(x, y) \, dx \, dy \\
 &= \frac{1}{4}
 \end{aligned}$$

Thus $P[XY \leq \frac{1}{2}] = \frac{3}{4}$.

Otherwise we would need to break it apart in two parts (when $y \leq \frac{1}{2}$ and when $y > \frac{1}{2}$):

$$\begin{aligned}
 P[XY \leq \frac{1}{2}] &= \int_{\frac{1}{2}}^1 \int_0^{\frac{1}{2y}} f(x, y) \, dx \, dy + \int_0^{\frac{1}{2}} \int_0^1 f(x, y) \, dx \, dy \\
 &= \frac{3}{4}
 \end{aligned}$$

5. We have

$$\begin{aligned}
 f_1(x) &= \int_{-\infty}^{\infty} f(x, y) \, dy \\
 &= \int_0^1 f(x, y) \, dy \\
 &= \int_0^1 x + y \, dy \\
 &= x + \frac{1}{2} \quad 0 < x < 1
 \end{aligned}
 \quad A = \{(x, y) \mid 0 < x < 1, 0 < y < 1\}$$

Similarly $f_2(y) = \int_0^1 f(x, y) \, dx = y + \frac{1}{2}$ for $0 < y < 1$.

6. If $x \leq 0$ or $y \leq 0$, then $F(x, y) = 0$.

Similarly if $x \geq 1$ and $y \geq 1$, then $F(x, y) = 1$.

If $0 < x \leq 1$ and $0 < y \leq 1$

$$\begin{aligned} F(x, y) &= \int_0^y \int_0^x f(x, y) \, dx \, dy \\ &= \frac{1}{2}x^2y + \frac{1}{2}xy^2 \end{aligned}$$

If $0 < x \leq 1$ and $y > 1$

$$\begin{aligned} F(x, y) &= P[X \leq x, Y \leq y] \\ &= P[X \leq x, Y \leq 1] \\ &= F(x, 1) \\ &= \frac{1}{2}(x^2 + x) \end{aligned}$$

Similarly for $x > 1$ and $0 < y \leq 1$, $F(x, y) = \frac{1}{2}(y^2 + y)$.

7. Note that $F_1(x) = \lim_{y \rightarrow \infty} F(x, y)$. From above we have

$$F_1(x) = \begin{cases} \lim_{y \rightarrow \infty} F(x, y) = \lim_{y \rightarrow \infty} 0 = 0 & x \leq 0 \\ \lim_{y \rightarrow \infty} F(x, y) = \lim_{y \rightarrow \infty} 1 = 1 & x \geq 1 \\ \lim_{y \rightarrow \infty} F(x, y) = \lim_{y \rightarrow \infty} \frac{1}{2}(x^2 + x) & 0 < x < 1 \end{cases}$$

6 September 21, 2018

6.1 Independence

X and Y are independent if $P[X \in A, Y \in B] = P[X \in A]P[Y \in B]$ for any $A, B \subseteq \mathbb{R}$.

Corollary 6.1. If X and Y are independent, then $h(X)$ and $g(Y)$ are independent for any real-valued functions $h(\cdot)$ and $g(\cdot)$.

Proof. To show $h(X), g(Y)$ are independent, we need to prove

$$P[h(X) \in A^*, g(Y) \in B^*] = P[h(X) \in A^*]P[g(Y) \in B^*]$$

for any $A^*, B^* \subseteq \mathbb{R}$.

Note that for functions $h : A \rightarrow A^*$ and $g : B \rightarrow B^*$, $x \in A \iff h(x) \in A^*$ and similarly $y \in B \iff g(y) \in B^*$.

Thus $P[h(X) \in A^*, g(Y) \in B^*] = P[X \in A, Y \in B] = P[X \in A]P[Y \in B]$ as X, Y are independent.

Again since we have an \iff correspondence we have $P[h(X) \in A^*]P[g(Y) \in B^*]$. □

Theorem 6.1. X, Y are independent **if and only if** either

$$f(x, y) = f_1(x)f_2(y) \quad \forall (x, y) \in A_1 \times A_2$$

where A_1, A_2 are the support sets for X and Y , respectively, OR

$$F(x, y) = F_1(x)F_2(y) \quad \forall (x, y) \in \mathbb{R} \times \mathbb{R}$$

Example 6.1. For $f(x, y) = x + y$, $0 < x < 1$, $0 < y < 1$, are X, Y independent? Why?

Note that from before we found that $f_1(x) = \frac{1}{2} + x$ for $0 < x < 1$; $f_2(y) = \frac{1}{2} + y$ for $0 < y < 1$.

Does $f(x, y) = f_1(x)f_2(y)$ for all $(x, y) \in A_1 \times A_2$, where $A_1 = \{x \mid 0 < x < 1\}$, $A_2 = \{y \mid 0 < y < 1\}$.
 No: since $(x + y) \neq (\frac{1}{2} + x)(\frac{1}{2} + y)$ for all $(x, y) \in (0, 1) \times (0, 1)$, thus X, Y are not independent.

6.2 Factorization independence theorem

Theorem 6.2 (Factorization independence). Suppose X, Y have joint pdf $f(x, y)$ and support set $A = \{(x, y) \mid f(x, y) > 0\}$.

Then X, Y are independent **if and only if** $A = A_1 \times A_2$ and $f(x, y) = h(x) \cdot g(y)$ for some non-negative functions $h(\cdot)$ and $g(\cdot)$ for all $(x, y) \in A$.

Remark 6.1. We need to check that

1. $A = A_1 \times A_2$ i.e. A is rectangular (otherwise we would have undefined values for $f(x, y)$ for some $x \in A_1$ or $y \in A_2$).
2. Check $f(x, y) = h(x) \cdot g(y)$

Example 6.2. Suppose X, Y have joint pdf

$$f(x, y) = \frac{\theta^{x+y} e^{-2\theta}}{x!y!} \quad x, y = 0, 1, 2, \dots$$

Are X, Y independent? Why?

1. Does $A = A_1 \times A_2$? Yes since we have $A_1 = \{x \mid x = 0, 1, 2, \dots\}$ and $A_2 = \{y \mid y = 0, 1, 2, \dots\}$.
2. We see that

$$f(x, y) = \left(\frac{\theta^x}{x!}\right) \left(\frac{\theta^y e^{-2\theta}}{y!}\right)$$

and there are many other functions where each function has complementary constant scaling factors.

Remark 6.2. Note that $h(\cdot)$ and $g(\cdot)$ may not be true pdfs (i.e. they may not sum up to 1 over the support set: see the remark below).

Thus X, Y are independent by the Factorization theorem.

Remark 6.3. When the Factorization theorem holds, $h(x)$ is *proportional* to $f_1(x)$ and $g(y)$ is proportional to $f_2(y)$.

Proof. We have

$$\begin{aligned} f_1(x) &= \sum_{y=0}^{\infty} f(x, y) \\ &= \sum_{y=0}^{\infty} h(x)g(y) \\ &= h(x) \sum_{y=0}^{\infty} g(y) \end{aligned}$$

From the example above, we had $g(y) = \frac{\theta^y}{y!}$, so

$$f_1(x) = h(x) \sum_{y=0}^{\infty} \frac{\theta^y}{y!} = e^{\theta} h(x) = \frac{\theta^x e^{-\theta}}{x!}$$

Thus $X \sim POI(\theta)$ and similarly $Y \sim POI(\theta)$. □

Example 6.3. Suppose X, Y have joint pdf

$$f(x, y) = \frac{2}{\pi} \quad 0 \leq x \leq \sqrt{1-y^2}, -1 \leq y \leq 1$$

Are X, Y independent? Why?

Note that $A \neq A_1 \times A_2$ since we have $A_1 = \{x \mid 0 \leq x \leq 1\}$ and $A_2 = \{y \mid -1 \leq y \leq 1\}$.

Since A does not have the support set that is the rectangular bounds of $A_1 \times A_2$ there is no way to factorize our joint pdf into the product of two marginal pdfs.

7 September 24, 2018

7.1 Conditional pmf/pdf

Definition 7.1. We define the **conditional pmf/pdf** of x on y to be

$$f_1(x \mid y) = \frac{f(x, y)}{f_2(y)} \quad (x, y) \in A \text{ and } f_2(y) \neq 0$$

where A is the support set for (X, Y) (i.e. $f(x, y)$)

Properties of $f_1(x \mid y)$ for discrete and continuous r.v's:

Discrete r.v.s 1. $f_1(x \mid y) \geq 0$ for all $(x, y) \in \mathbb{R} \times \mathbb{R}$

$$2. \sum_{x \in \mathbb{R}} f_1(x \mid y) = 1$$

Continuous r.v.s 1. $f_1(x \mid y) \geq 0$ for all $(x, y) \in \mathbb{R} \times \mathbb{R}$

$$2. \int_{-\infty}^{\infty} f_1(x \mid y) dx = 1$$

Similarly $f_2(y \mid x) = \frac{f(x, y)}{f_1(x)}$ and $f_1(x) \neq 0$.

7.2 Product rule

The product rule states that

$$\begin{aligned} f(x, y) &= f_1(x \mid y) f_2(y) \\ &= f_2(y \mid x) f_1(x) \end{aligned} \quad \text{OR}$$

Application of product rule: if $f_1(x \mid y)$ and $f_2(y)$ are given, we can find $f_1(x)$ (Take $\int_{y \in A} f_1(x \mid y) f_2(y) dy$ in the continuous case).

Example 7.1. Let $Y \sim POI(\mu)$ and $X \mid Y = y \sim BIN(y, p)$. Find the marginal distribution of X .

We will take the route $f_1(x \mid y)$ and $f_2(y) \rightarrow f(x, y) \rightarrow f_1(x)$.

Note that

$$f_2(y) = \frac{\mu^y e^{-\mu}}{y!} \quad y = 0, 1, 2, \dots$$

Also

$$f_1(x | y) = \binom{y}{x} p^x (1-p)^{y-x} \quad x = 0, 1, \dots, y$$

Thus we have

$$\begin{aligned} f(x, y) &= f_1(x | y) f_2(y) = \frac{\mu^y e^{-\mu}}{y!} \cdot \frac{y!}{(y-x)!x!} p^x (1-p)^{y-x} \\ &= \frac{\mu^y e^{-\mu}}{(y-x)!x!} p^x (1-p)^{y-x} \end{aligned}$$

where $x = 0, 1, \dots, y$ and $y = 0, 1, \dots$ i.e. $0 \leq x \leq y$ (and $y \geq 0$). We need to be aware of these bounds when marginalizing over x , so

$$\begin{aligned} f_1(x) &= \sum_{y=x}^{\infty} f(x, y) \\ &= \frac{e^{-\mu} p^x (1-p)^{-x}}{x!} \sum_{y=x}^{\infty} \frac{\mu^y (1-p)^y}{(y-x)!} \\ &= \frac{e^{-\mu} \left(\frac{p}{1-p}\right)^x}{x!} \sum_{y=x}^{\infty} \frac{(\mu(1-p))^y}{(y-x)!} \\ &= \frac{e^{-\mu} \left(\frac{p}{1-p}\right)^x (\mu(1-p))^x}{x!} \sum_{y=x}^{\infty} \frac{(\mu(1-p))^{y-x}}{(y-x)!} \\ &= \frac{e^{-\mu} (\mu p)^x}{x!} \sum_{n=0}^{\infty} \frac{(\mu(1-p))^n}{n!} \quad n = y - x \\ &= \frac{e^{-\mu} (\mu p)^x}{x!} e^{\mu(1-p)} \quad \text{Taylor series of } e^{\mu(1-p)} \\ &= \frac{e^{-\mu p} (\mu p)^x}{x!} \quad x = 0, 1, \dots \end{aligned}$$

that is $X \sim POI(\mu p)$.

Example 7.2. let $Y \sim GAM(\alpha, 1)$ (not $GAM(\alpha, \frac{1}{\theta})$ in the notes) and $X | Y = y \sim WEI(y^{-\frac{1}{p}}, p)$ (Weibull distribution). Find the marginal pdf of X .

We will be following $f_1(x | y)$ and $f_2(y) \rightarrow f(x, y) \rightarrow f_1(x)$.

Note that

$$f_1(y) = \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y}$$

For a given $X \sim WEI(\theta, \beta)$ we have

$$f(x) = \frac{\beta}{\theta \beta} x^{\beta-1} e^{-(\frac{x}{\theta})^\beta}$$

where $x > 0$. Thus we have

$$f_1(x | y) = \frac{p}{(y^{-\frac{1}{p}})^p} x^{p-1} e^{-\left(\frac{x}{y^{-\frac{1}{p}}}\right)^p}$$

Note we have $A = \{(x, y) \mid x > 0, y > 0\}$ thus

$$\begin{aligned} f_1(x) &= \int_0^\infty f(x, y) \, dy \\ &= \frac{px^{p-1}}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y} y e^{-x^p y} \, dy \\ &= \frac{px^{p-1}}{\Gamma(\alpha)} \int_0^\infty y^\alpha e^{-y(1+x^p)} \, dy \end{aligned}$$

Recall we have

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

So we have

$$\int_0^\infty y^\alpha e^{-y(1+x^p)} \, dy = \Gamma(\alpha+1)[(1+x^p)^{-1}]^{\alpha+1}$$

Thus

$$\begin{aligned} f_1(x) &= \frac{px^{p-1}}{\Gamma(\alpha)} \Gamma(\alpha+1)[(1+x^p)^{-1}]^{\alpha+1} \\ &= p\alpha \cdot \frac{x^{p-1}}{(1+x^p)^{\alpha+1}} \end{aligned}$$

Note that $X \sim \text{Burr}(p, \alpha)$ or the Burr distribution.

7.3 Independence and condition pmf/pdfs

Note that $f(x, y) = f_1(x)f_2(y) = f_1(x \mid y)f_2(y)$, therefore X and Y are independent **if and only if** $f_1(x \mid y) = f_1(x)$ (or similarly if $f_2(y \mid x) = f_2(y)$).

Example 7.3. Let $f(x, y) = \frac{2}{\pi}$ where $0 \leq x \leq \sqrt{1-y^2}$, $-1 \leq y \leq 1$.

Note that

$$\begin{aligned} f_1(x) &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f(x, y) \, dy = \frac{4\sqrt{1-x^2}}{\pi} & 0 < x < 1 \\ f_2(y) &= \int_0^{\sqrt{1-y^2}} f(x, y) \, dx = \frac{2\sqrt{1-y^2}}{\pi} & -1 < y < 1 \\ f_1(x \mid y) &= \frac{f(x, y)}{f_2(y)} = \frac{1}{\sqrt{1-y^2}} & 0 \leq x \leq \sqrt{1-y^2}, -1 < y < 1 \end{aligned}$$

Since $f_1(x, y) \neq f_1(x)$ then X and Y are not independent.

8 September 26, 2018

8.1 Joint expectation

We define the **joint expectation** for discrete and continuous r.v.s:

Discrete The joint expectation is

$$E[h(X, Y)] = \sum_{x \in \mathbb{R}} \sum_{y \in \mathbb{R}} h(x, y) \cdot f(x, y)$$

Continuous The joint expectation is

$$E[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y) \, dx \, dy$$

Theorem 8.1. If X, Y are independent, then

$$E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)]$$

for any real-valued functions $g(\cdot)$ and $h(\cdot)$.

Proof. Note that $g(X)$ and $h(Y)$ are functions of X and Y , thus by the joint expectation

$$\begin{aligned} E[g(X)h(Y)] &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(x)h(y)f(x, y) \, dx \right] dy \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(x)h(y)f_1(x)f_2(y) \, dx \right] dy && \text{independence} \\ &= \left[\int_{-\infty}^{\infty} g(x)f_1(x) \, dx \right] \left[\int_{-\infty}^{\infty} h(y)f_2(y) \, dy \right] \\ &= E[g(X)] \cdot E[h(Y)] \end{aligned}$$

□

8.2 Conditional expectation

We define the **conditional expectation** for discrete and continuous r.v.s:

Discrete The conditional expectation is

$$E[Y \mid X = x] = \sum_{y \in \mathbb{R}} y f_2(y \mid x)$$

by LOTUS

$$E[h(Y) \mid X = x] = \sum_{y \in \mathbb{R}} h(y) f_2(y \mid x)$$

Continuous The conditional expectation is

$$E[Y \mid X = x] = \int_{-\infty}^{\infty} y f_2(y \mid x) \, dy$$

by LOTUS

$$E[h(Y) | X = x] = \int_{-\infty}^{\infty} h(y) f_2(y | x) dy$$

Remark 8.1. 1. $E[Y | X = x]$ is a function of x only since we've summed over our support for Y .

2. If X, Y are independent, then $E[Y | X = x] = E[Y]$ since

$$\begin{aligned} E[Y | X = x] &= \int_{-\infty}^{\infty} y f_2(y | x) dy \\ &= \int_{-\infty}^{\infty} y f_2(y) dy && \text{independence} \\ &= E[Y] \end{aligned}$$

similarly $E[h(Y) | X = x] = E[h(Y)]$.

Example 8.1. Let $f(x, y) = \frac{2}{\pi}$ where $0 \leq x \leq \sqrt{1 - y^2}$, $-1 \leq y \leq 1$.

Note that $A = \{(x, y) | 0 \leq x \leq \sqrt{1 - y^2}, -1 \leq y \leq 1\}$ or $A = \{(x, y) | 0 \leq x \leq 1, -\sqrt{1 - x^2} \leq y \leq \sqrt{1 - x^2}\}$, where $A_1 = \{x | 0 \leq x \leq 1\}$ and $A_2 = \{y | -1 \leq y \leq 1\}$.

Thus the conditional pdfs are

$$f_2(y | x) = \frac{f(x, y)}{f_1(x)} = \frac{1}{2\sqrt{1 - x^2}}$$

for $(x, y) \in A$ and $f_1(x) \neq 0$ thus $0 \leq x < 1$ and $-\sqrt{1 - x^2} \leq y \leq \sqrt{1 - x^2}$.

Note that $Y | X = x$ is actually a uniform distribution symmetric around $y = 0$ ($UNIF(-\sqrt{1 - x^2}, \sqrt{1 - x^2})$ for $0 \leq x < 1$), thus we expect $E[Y | X = x] = 0$. We verify

$$\begin{aligned} E[Y | X = x] &= \int_{-\sqrt{1 - x^2}}^{\sqrt{1 - x^2}} y f_2(y | x) dy \\ &= \frac{1}{2\sqrt{1 - x^2}} \left(\frac{1}{2} y^2 \Big|_{-\sqrt{1 - x^2}}^{\sqrt{1 - x^2}} \right) \\ &= \frac{1}{2\sqrt{1 - x^2}} \cdot 0 \\ &= 0 \end{aligned}$$

We can also find $E[Y^2 | X = x]$

$$\begin{aligned} E[Y^2 | X = x] &= \int_{-\sqrt{1 - x^2}}^{\sqrt{1 - x^2}} y^2 f_2(y | x) dy \\ &= \frac{1}{2\sqrt{1 - x^2}} \left(\frac{1}{3} y^3 \Big|_{-\sqrt{1 - x^2}}^{\sqrt{1 - x^2}} \right) \\ &= \frac{(1 - x^2)}{3} \quad 0 \leq x < 1 \end{aligned}$$

For $Var(Y | X = x)$ we have

$$\begin{aligned} Var(Y | X = x) &= E[Y^2 | X = x] - E[Y | X = x]^2 \\ &= \frac{(1 - x^2)}{3} - 0^2 \\ &= \frac{(1 - x^2)}{3} \quad 0 \leq x < 1 \end{aligned}$$

Remark 8.2. $E[Y | X = x]$ and $E[h(Y) | X = x]$ are functions of x , thus $E[Y | X]$ is a function of X (function of a random variable is a random variable).

8.3 Expectation of a conditional expectation

Theorem 8.2. We claim $E[E[h(Y) | X]] = E[h(Y)]$.

Let $g(X) = E[h(Y) | X]$, thus we have a function of X which from LOTUS we know

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{\infty} g(x) f_1(x) dx \\ &= \int_{-\infty}^{\infty} E[h(Y) | X = x] f_1(x) dx && g(x) = E[h(Y) | X = x] \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} h(y) f_2(y | x) dy \right] f_1(x) dx \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} h(y) f_2(y | x) f_1(x) dy \right] dx \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} h(y) f(x, y) dy \right] dx \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} h(y) f(x, y) dx \right] dy && \text{Fubini's theorem} \\ &= \int_{-\infty}^{\infty} h(y) \left[\int_{-\infty}^{\infty} f(x, y) dx \right] dy \\ &= \int_{-\infty}^{\infty} h(y) f_2(y) dy \\ &= E[h(Y)] \end{aligned}$$

8.4 Variance as sum of conditional expectations

Theorem 8.3. We claim $Var(Y) = E[Var(Y | X)] + Var(E[Y | X])$.

We have $Var(Y) = E[Y^2] - E[Y]^2$ on the LHS.

On the RHS we have

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

where the last equality follows from $E[E[h(Y) | X]] = E[h(Y)]$.

Example 8.2. Suppose $Y \mid P = p \sim \text{BIN}(n, p)$ and $P \sim \text{UNIF}(0, 1)$. Find $E[Y]$ and $\text{Var}(Y)$.

Note that

$$E[Y] = E[E[Y \mid P]] = E[nP] = n \cdot E[P] = \frac{n}{2}$$

Similarly

$$\begin{aligned} \text{Var}(Y) &= E[\text{Var}(Y \mid P)] + \text{Var}(E[Y \mid P]) = E[nP(1 - P)] + \text{Var}(nP) \\ &= nE[P] - nE[P^2] + n^2\text{Var}(P) \\ &= n\frac{1}{2} - n(\text{Var}(P) + E[P]^2) + n^2\text{Var}(P) \\ &= \frac{n}{2} - n\left(\frac{1}{12} + \frac{1}{2^2}\right) + n^2\frac{1}{12} \\ &= \frac{5n}{6} + \frac{n^2}{12} \end{aligned}$$

9 September 28, 2018

9.1 Joint moment generating function (mgf)

Recall the moment generating function (mgf) of X is defined as $M_X(t) = E[e^{tX}]$. For a given MGF:

1. State the values of t such that $M_X(t)$ exists, i.e. $E[e^{tX}] < \infty$.
2. Uniqueness: if X and Y have the same mgf, then X and Y are identically distributed (i.e. X, Y have the same pmf/pdf, cdf, etc.).

Definition 9.1 (Joint mgf). The **joint mgf** of X and Y is defined as

$$M(t_1, t_2) = E[e^{(t_1, t_2) \cdot (X, Y)^T}] = E[e^{t_1 X + t_2 Y}] = E[e^{t_1 X} e^{t_2 Y}]$$

where one needs to state the values of t_1, t_2 such that $M(t_1, t_2)$ exists.

Note that given the joint mgf, it is straightforward to derive the marginal mgf

$$\begin{aligned} M_X(t_1) &= M(t_1, 0) = E[e^{t_1 X}] \\ M_Y(t_2) &= M(0, t_2) = E[e^{t_2 Y}] \end{aligned}$$

Example 9.1. Suppose $f(x, y) = e^{-y}$, $0 < x < y$. Find $M(t_1, t_2)$ and $M_X(t)$.

$$\begin{aligned} M(t_1, t_2) &= E[e^{t_1 X + t_2 Y}] = \int_0^\infty \left(\int_0^y e^{t_1 x} e^{t_2 y} f(x, y) dx \right) dy \\ &= \int_0^\infty e^{(t_2 - 1)y} \left(\int_0^y e^{t_1 x} dx \right) dy \\ &= \frac{1}{t_1} \int_0^\infty e^{(t_2 - 1)y} e^{t_1 y} - 1 dy \\ &= \frac{1}{t_1} \int_0^\infty e^{(t_2 - 1)y} \cdot (e^{t_1 y} - 1) dy \\ &= \frac{1}{t_1} \left(\int_0^\infty e^{(t_1 + t_2 - 1)y} dy - \int_0^\infty e^{(t_2 - 1)y} dy \right) \end{aligned}$$

Note that

$$\int_0^\infty e^{(t_1+t_2-1)y} dy = \frac{1}{t_1+t_2-1} e^{(t_1+t_2-1)y} \Big|_0^\infty$$

Taking the limit

$$\lim_{y \rightarrow \infty} e^{(t_1+t_2-1)y} < \infty = 0$$

iff $t_1 + t_2 - 1 < 0$. Similarly $t_2 - 1 < 0$ must hold from our other integral.

So we have

$$\begin{aligned} M(t_1, t_2) &= \frac{1}{t_1} \left(\frac{1}{t_1+t_2-1} (0-1) + \frac{1}{t_2-1} (0-1) \right) \\ &= \frac{1}{(t_2-1)(t_1+t_2-1)} \end{aligned}$$

For $M_X(t)$ we have

$$M_X(t) = M(t, 0) = \frac{1}{1-t}$$

where $t < 1$ (from our two constraints on t_1, t_2).

For $M_Y(t)$ we have

$$M_Y(t) = M(0, t) = \frac{1}{(1-t)^2}$$

where $t_2 < 1$ (from our two constraints on t_1, t_2).

Recall $X \sim \text{GAM}(\alpha, \beta)$ has $M_X(t) = \frac{1}{(1-\beta t)^\alpha}$, $t < \frac{1}{\beta}$.

Due to the uniqueness of mgf, $X \sim \text{GAM}(1, 1)$ and $Y \sim \text{GAM}(2, 1)$.

9.2 Independence and joint mgfs

X and Y are independent if and only if

$$M(t_1, t_2) = M_X(t_1)M_Y(t_2) \quad \forall t_1 \in B_1, \forall t_2 \in B_2$$

where $M_X(t_1)$ exists in B_1 and $M_Y(t_2)$ exists in B_2 (i.e. the bounds on t_1, t_2 such that $M(t_1, t_2)$ is well-defined).

Example 9.2. From our previous example where

$$M(t_1, t_2) = \frac{1}{(t_2-1)(t_1+t_2-1)}$$

and $M_X(t) = \frac{1}{1-t}$ and $M_Y(t) = \frac{1}{(1-t)^2}$, clearly $M(t_1, t_2) \neq M_X(t_1)M_Y(t_2)$ so X, Y are not independent.

9.3 Summary of methods for verifying independence

The following are equivalent (TFAE) for showing independence of two r.v.s X, Y :

joint pmf/pdf Show $f(x, y) = f_1(x)f_2(y)$

joint cdf Show $F(x, y) = F_1(x)F_2(y)$

Factorization Theorem Show $f(x, y) = h(x)g(y)$ and support set is the rectangular Cartesian product of the individual support sets.

conditional pdf Show $f_1(x | y) = f_1(x)$.

joint mgf Show $M(t_1, t_2) = M_X(t_1)M_Y(t_2)$ (for all $(t_1, t_2) \in B$).

10 October 1, 2018

10.1 Expectations/moments from mgf

Suppose X and Y have joint mgf $M(t_1, t_2)$ for all $t_1 \in (-h_1, h_1), t_2 \in (-h_2, h_2)$, some $h_1, h_2 > 0$. Find $E[XY^2]$ and $E[X^k Y^j]$, $k, j = 0, 1, 2, \dots$

Proof. We can use the moment generating functions to find the expectations.

In the continuous case

$$M(t_1, t_2) = \int \left(\int e^{t_1 x} e^{t_2 y} f(x, y) dx \right) dy$$

Thus we have

$$\begin{aligned} \frac{\partial M(t_1, t_2)}{\partial t_1} &= \frac{\partial}{\partial t_1} \int \left(\int e^{t_1 x} e^{t_2 y} f(x, y) dx \right) dy \\ &= \int \left(\int \left(\frac{\partial}{\partial t_1} e^{t_1 x} \right) e^{t_2 y} f(x, y) dx \right) dy \\ &= \int \left(\int x e^{t_1 x} e^{t_2 y} f(x, y) dx \right) dy \end{aligned}$$

We then take

$$\begin{aligned} \frac{\partial^2}{\partial t_1 \partial t_2} M(t_1, t_2) &= \frac{\partial}{\partial t_2} \left(\frac{\partial}{\partial t_1} M(t_1, t_2) \right) \\ &= \frac{\partial}{\partial t_2} \int \left(\int x e^{t_1 x} e^{t_2 y} f(x, y) dx \right) dy \\ &= \int \left(\int x e^{t_1 x} \left(\frac{\partial}{\partial t_2} e^{t_2 y} \right) f(x, y) dx \right) dy \\ &= \int \left(\int x e^{t_1 x} y e^{t_2 y} f(x, y) dx \right) dy \end{aligned}$$

Once more

$$\begin{aligned} \frac{\partial^3}{\partial t_1 \partial t_2^2} M(t_1, t_2) &= \frac{\partial}{\partial t_2} \left(\frac{\partial^2}{\partial t_1 \partial t_2} M(t_1, t_2) \right) \\ &= \int \left(\int x e^{t_1 x} y^2 e^{t_2 y} f(x, y) dx \right) dy \end{aligned}$$

Thus if we continue in this fashion

$$\frac{\partial^{k+j}}{\partial t_1^k \partial t_2^j} M(t_1, t_2) = \int \left(\int x^k e^{t_1 x} y^j e^{t_2 y} f(x, y) dx \right) dy$$

To find $E[XY^2]$, we simply let $t_1 = t_2 = 0$ in $\frac{\partial^2}{\partial t_1 \partial t_2} M(t_1, t_2)$

$$\begin{aligned} \left(\frac{\partial^3}{\partial t_1 \partial t_2^2} M(t_1, t_2) \right) \big|_{t_1=t_2=0} &= \int \left(\int x e^{0x} y^2 e^{0y} f(x, y) dx \right) dy \\ &= \int \left(\int x y^2 f(x, y) dx \right) dy \\ &= E[XY^2] \end{aligned}$$

Similarly

$$\left(\frac{\partial^{k+j}}{\partial t_1^k \partial t_2^j} M(t_1, t_2) \right) \big|_{t_1=t_2=0} = E[X^k Y^j]$$

This also holds for $E[X^k]$ where

$$\left(\frac{\partial^k}{\partial t_1^k} M(t_1, t_2) \right) \big|_{t_1=t_2=0} = E[X^k]$$

i.e. $j = 0$. □

10.2 Multinomial distribution

Definition 10.1 (Multinomial distribution). Let $(X_1, X_2, \dots, X_k) \sim MULT(n, p_1, p_2, \dots, p_k)$ where

$$f(x_1, x_2, \dots, x_k) = \frac{n!}{x_1! x_2! \dots x_k! (n - \sum_{i=1}^k x_i)!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k} (1 - \sum_{i=1}^k p_i)^{n - \sum_{i=1}^k x_i}$$

where $0 \leq x_i \leq n$, $0 \leq \sum_{i=1}^k x_i \leq n$, $0 \leq p_i \leq 1$, $0 \leq \sum_{i=1}^k p_i \leq 1$.

Remark 10.1. The k random variables represents a random sample of size n where each unit in this random sample could be one of $k + 1$ types with corresponding probabilities $p_1, p_2, \dots, p_k, 1 - \sum_{i=1}^k p_i$ and x_i is the number elements of the i th type.

Remark 10.2. Binomial $BIN(n, p)$ is a special case of $MULT$ i.e. there are 2 types with probabilities p and $1 - p$ i.e. $MULT(n, p)$ with $k = 1$.

Exercise 10.1 (Hardy-Weinberg law of genetics). We have a random sample of size n from the population. Each unit/person in this sample could be one of 3 genotypes: “AA” with probability $p_1 = \theta^2$, “Aa” with $p_2 = 2\theta(1 - \theta)$, and “aa” with probability $p_3 = (1 - \theta)^2$, $0 < \theta < 1$ i.e. $0 < p_i < 1$ and $\sum_{i=1}^3 p_i = 1$.

Let X_1, X_2 be the number of type “AA” and “Aa”, respectively.

Thus

$$P[X_1 = x_1, X_2 = x_2] = \frac{n!}{x_1! x_2! (n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2}$$

where $0 \leq x_1, x_2 \leq n$ and $0 \leq x_1 + x_2 \leq n$ i.e. $(X_1, X_2) \sim MULT(n, p_1, p_2)$.

10.3 Mgf of multinomial distribution

Note that the MGF for $MULT(n, p_1, p_2)$

$$\begin{aligned}
 M(t_1, t_2) &= E[e^{t_1 X_1 + t_2 X_2}] \\
 &= \sum_{x_1=0}^n \sum_{x_2=0}^{n-x_1} e^{t_1 x_1} e^{t_2 x_2} f(x_1, x_2) \\
 &= \sum_{x_1=0}^n \sum_{x_2=0}^{n-x_1} e^{t_1 x_1} e^{t_2 x_2} \frac{n!}{x_1! x_2! (n-x_1-x_2)!} p_1^{x_1} p_2^{x_2} (1-p_1-p_2)^{n-x_1-x_2}
 \end{aligned}$$

Recall the Multinomial series identity where

$$(a+b+c)^n = \sum_{x_1=0}^n \sum_{x_2=0}^{n-x_1} \frac{n!}{x_1! x_2! (n-x_1-x_2)!} a^{x_1} b^{x_2} c^{n-x_1-x_2}$$

for any $a, b, c \in \mathbb{R}$. Thus we have

$$M(t_1, t_2) = (e^{t_1} p_1 + e^{t_2} p_2 + 1 - p_1 - p_2)^n$$

For $e^{t_1} p_1, e^{t_2} p_2 \in \mathbb{R}$, we require $t_1, t_2 \in \mathbb{R}$.

In general for $MULT(n, p_1, \dots, p_k)$

$$M(t_1, \dots, t_k) = (e^{t_1} p_1 + \dots + e^{t_k} p_k + 1 - \sum_{i=1}^k p_i)^n$$

10.4 Subset of multinomial is multinomial

Claim. Any “subset” of a multinomial still has a multinomial distribution.

For example suppose we had $(X_1, \dots, X_6) \sim MULT(n, p_1, \dots, p_6)$. We have $(X_1, X_3, X_5) \sim MULT(n, p_1, p_3, p_5)$.

Proof. Note that $M(t_1, \dots, t_6) = (e^{t_1} p_1 + \dots + e^{t_6} p_6 + 1 - \sum_{i=1}^6 p_i)^n$, thus

$$\begin{aligned}
 M_{X_1, X_3, X_5}(t_1, t_3, t_5) &= E[e^{t_1 X_1 + t_3 X_3 + t_5 X_5}] \\
 &= E[e^{t_1 X_1 + 0 X_2 + t_3 X_3 + 0 X_4 + t_5 X_5 + 0 X_6}] \\
 &= M(t_1, t_2 = 0, t_3, t_4 = 0, t_5, t_6 = 0) \\
 &= (e^{t_1} p_1 + e^0 p_2 + e^{t_3} p_3 + e^0 p_4 + e^{t_5} p_5 + e^0 p_6 + 1 - \sum_{i=1}^6 p_i)^n \\
 &= (e^{t_1} p_1 + p_2 + e^{t_3} p_3 + p_4 + e^{t_5} p_5 + p_6 + 1 - \sum_{i=1}^6 p_i)^n \\
 &= (e^{t_1} p_1 + e^{t_3} p_3 + e^{t_5} p_5 + 1 - p_1 - p_3 - p_5)^n
 \end{aligned}$$

which is the mgf of $MULT(n, p_1, p_3, p_5)$. By the uniqueness of mgfs our claim holds. \square

11 October 3, 2018

11.1 More multinomial problems

Example 11.1. Let $T = x_i + x_j$, $1 \leq i \leq j \leq k$.

Claim. We claim $T \sim \text{BIN}(np_i + p_j)$.

Proof. For $(x_i, x_j) \sim \text{MULT}(n, p_i, p_j)$ we have the mgf $M(t_i, t_j) = (e^{t_i}p_i + e^{t_j}p_j + 1 - p_i - p_j)^n$ for all $t_i, t_j \in \mathbb{R}$. The mgf of T is $M_T(t) = E[e^{tT}] = E[e^{t(X_i + X_j)}] = E[e^{tX_i + tX_j}]$.

Thus

$$\begin{aligned} M_T(t) &= M(t_i = t, t_j = t) \\ &= (e^t p_i + e^t p_j + 1 - p_i - p_j)^n \\ &= (e^t (p_i + p_j) + 1 - (p_i + p_j))^n \end{aligned}$$

Recall the mgf of $X \sim \text{BIN}(n, p)$ is $M_X(t) = (e^t p + 1 - p)^n$ for all $t \in \mathbb{R}$, thus $T = x_i + x_j \sim \text{BIN}(n, p_i + p_j)$ by uniqueness of mgf. \square

Claim. We claim $\text{Cov}(X_i, X_j) = -np_i p_j$.

Proof. Note that $\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$. Also

$$E(X_i X_j) = \left(\frac{\partial^2}{\partial t_i \partial t_j} M(t_i, t_j) \right) \Big|_{t_i=t_j=0}$$

We have

$$\begin{aligned} \frac{\partial}{\partial t_i} M(t_i, t_j) &= \frac{\partial}{\partial t_i} (e^{t_i} p_i + e^{t_j} p_j + 1 - p_i - p_j)^n \\ &= n(e^{t_i} p_i + e^{t_j} p_j + 1 - p_i - p_j)^{n-1} \cdot e^{t_i} p_i \end{aligned}$$

Furthermore

$$\begin{aligned} \frac{\partial^2}{\partial t_i \partial t_j} M(t_i, t_j) &= \frac{\partial}{\partial t_j} \left(\frac{\partial}{\partial t_i} M(t_i, t_j) \right) \\ &= n(n-1)(e^{t_i} p_i + e^{t_j} p_j + 1 - p_i - p_j)^{n-2} \cdot e^{t_i} p_i \cdot e^{t_j} p_j \end{aligned}$$

Therefore

$$\begin{aligned} E(X_i X_j) &= n(n-1)(e^0 p_i + e^0 p_j + 1 - p_i - p_j)^{n-2} e^0 p_i e^0 p_j \\ &= n(n-1)p_i p_j \end{aligned}$$

So we have

$$\text{Cov}(X_i, X_j) = n(n-1)p_i p_j - (np_i)(np_j) = -np_i p_j$$

as $X_i \sim \text{BIN}(n, p_i)$ and $E(X_i) = np_i$. \square

Claim. We claim $(X_i \mid X_j = x_j) \sim \text{BIN}(n - x_j, \frac{p_i}{1-p_j})$.

Note that $(X_i, X_j) \sim \text{MULT}(n, p_i, p_j)$ and $X_j \sim \text{BIN}(n, p_j)$, thus

$$\begin{aligned}
 f(x_i | x_j) &= \frac{f(x_i, x_j)}{f(x_j)} \\
 &= \frac{\frac{n!}{x_i!x_j!(n-x_i-x_j)!} p_i^{x_i} p_j^{x_j} (1-p_i-p_j)^{n-x_i-x_j}}{\frac{n!}{(n-x_j)!x_j!} p_j^{x_j} (1-p_j)^{n-x_j}} \\
 &= \frac{(n-x_j)!}{(n-x_j-x_i)!x_i!} \frac{p_i^{x_i} (1-p_i-p_j)^{n-x_i-x_j}}{(1-p_j)^{n-x_j}} \\
 &= \frac{(n-x_j)!}{(n-x_j-x_i)!x_i!} \frac{p_i^{x_i}}{(1-p_j)^{x_i}} \frac{(1-p_i-p_j)^{n-x_i-x_j}}{(1-p_j)^{n-x_j-x_i}} \\
 &= \frac{(n-x_j)!}{(n-x_j-x_i)!x_i!} \left(\frac{p_i}{1-p_j}\right)^{x_i} \left(\frac{1-p_i-p_j}{1-p_j}\right)^{n-x_i-x_j} \\
 &= \frac{(n-x_j)!}{(n-x_j-x_i)!x_i!} \left(\frac{p_i}{1-p_j}\right)^{x_i} \left(1 - \frac{p_i}{1-p_j}\right)^{n-x_i-x_j}
 \end{aligned}$$

i.e. $f_i(x_i | x_j)$ is the same as the pmf of $\text{BIN}(n-x_j, \frac{p_i}{1-p_j})$ so the claim holds.

Claim. We claim $X_i | X_i + X_j = t \sim \text{BIN}(t, \frac{p_i}{p_i+p_j})$.

Proof. Note that

$$P(X_i = x_i, X_i + X_j = t) = P(X_i = x_i, X_j = t - x_i)$$

which is just our joint pmf.

Also from before we have $P(X_i + X_j = t)$ is the pmf of $T = X_i + X_j$ and $T \sim \text{BIN}(n, p_i + p_j)$.

Thus we have

$$\begin{aligned}
 f_i(x_i | t) &= \frac{P(X_i = x_i, X_j = t - x_i)}{P(T = t)} \\
 &= \frac{\frac{n!}{x_i!(t-x_i)!(n-t)!} p_i^{x_i} p_j^{t-x_i} (1-p_i-p_j)^{n-t}}{\frac{n!}{t!(n-t)!} (p_i + p_j)^t (1-p_i-p_j)^{n-t}} \\
 &= \frac{t!}{x_i!(t-x_i)!} \frac{p_i^{x_i} p_j^{t-x_i}}{(p_i + p_j)^t} \\
 &= \frac{t!}{x_i!(t-x_i)!} \left(\frac{p_i}{p_i + p_j}\right)^{x_i} \left(1 - \frac{p_i}{p_i + p_j}\right)^{t-x_i}
 \end{aligned}$$

which is the pmf of $\text{BIN}(t, \frac{p_i}{p_i+p_j})$. □

11.2 Bivariate normal distribution

Recall for a univariate normal distribution $X \sim N(\mu, \sigma^2)$

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

The bivariate normal distribution for $\vec{x} = (x_1, x_2)^T$ is denoted as $X \sim BVN(\vec{\mu}, \Sigma)$ where $\vec{\mu} = (E(X_1), E(X_2))^T$ and

$$\Sigma = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{bmatrix}$$

Notice that $Cov(X_1, X_2) = Cov(X_2, X_1)$ i.e. Σ is symmetric and positive definite.

We define the pdf for the bivariate normal distribution as

$$f(x_1, x_2) = \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1}(\vec{x}-\vec{\mu})}$$

for all $x_1, x_2 \in \mathbb{R}$.

12 October 5, 2018

12.1 Remarks of bivariate normal

Remark 12.1. 1.

$$\vec{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} E[X_1] \\ E[X_2] \end{bmatrix}$$

2.

$$\Sigma = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & P\sigma_1\sigma_2 \\ P\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

where

$$P = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)}\sqrt{Var(X_2)}}$$

and $-1 < P < 1$ (note if $P = \pm 1$ then Σ is not full rank thus Σ^{-1} does not exist).

3.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2 = 1$$

(useful for finding $M(t_1, t_2)$).

4. Σ is positive definite (symmetric by definition) i.e. for all

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

we have $\vec{y}^T \Sigma \vec{y} > 0$.

That is: both eigenvalues of Σ are positive.

Remark 12.2. If $X \sim BVN(\mu + \Sigma t, \Sigma)$ where $t = [t_1, t_2]^T$ then the joint pdf

$$g(x_1, x_2) = \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu-\Sigma t)^T \Sigma^{-1}(x-\mu-\Sigma t)}$$

for $x \in \mathbb{R}^2$ then

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) dx_1 dx_2 = 1$$

12.2 Mgf of bivariate normal

Claim. For $BVN(\mu, \Sigma)$ we claim its mgf is

$$M(t_1, t_2) = e^{t^T \mu + \frac{1}{2} t^T \Sigma t}$$

where $t = [t_1, t_2]^T \in \mathbb{R}^2$.

Proof. Note that the following properties hold in general

1. $a^T b = b^T a$ for $a, b \in \mathbb{R}^2$.
2. $\Sigma = \Sigma^T$ since Σ is symmetric.
3. $\Sigma \Sigma^{-1} = \Sigma^{-1} \Sigma = I_{2 \times 2}$
4. $Ia = a$ and $a^T I = a^T$.
5. $(\Sigma t)^T = t^T \Sigma^T$

We have

$$M(t_1, t_2) = E[e^{t^T x}]$$

where $t = [t_1, t_2]^T, x = [x_1, x_2]^T$.

$$\begin{aligned} M(t_1, t_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t^T x} \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}((x-\mu)^T \Sigma^{-1}(x-\mu) - 2t^T x)} dx_1 dx_2 \end{aligned}$$

If we can show that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) - 2t^T x} dx_1 dx_2 = e^{t^T \mu + \frac{1}{2} t^T \Sigma t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu^*)^T \Sigma^{-1}(x-\mu^*)} dx_1 dx_2$$

for some μ^* then we are done since the integral on the right is just the total probability of a bivariate r.v. $N(\mu^*, \Sigma)$ which is 1.

$$\begin{aligned} &e^{t^T \mu + \frac{1}{2} t^T \Sigma t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu^*)^T \Sigma^{-1}(x-\mu^*)} dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}((x-\mu^*)^T \Sigma^{-1}(x-\mu^*) - 2(t^T \mu + \frac{1}{2} t^T \Sigma t))} dx_1 dx_2 \end{aligned}$$

We notice that every term is the same except for the exponential terms: thus we need to show the exponential terms are equivalent, specifically the terms in the exponent:

$$(x - \mu)^T \Sigma^{-1}(x - \mu) - 2t^T x = (x - \mu^*)^T \Sigma^{-1}(x - \mu^*) - 2(t^T \mu + \frac{1}{2} t^T \Sigma t)$$

We claim $\mu^* = \mu + \Sigma t$. We have on the RHS

$$\begin{aligned}
& (x - \mu - \Sigma t)^T \Sigma^{-1} (x - \mu - \Sigma t) - 2(t^T \mu + \frac{1}{2} t^T \Sigma t) \\
&= ((x - \mu)^T - (\Sigma t)^T) \Sigma^{-1} ((x - \mu) - (\Sigma t)) - 2(t^T \mu + \frac{1}{2} t^T \Sigma t) \\
&= (x - \mu)^T \Sigma^{-1} (x - \mu) + (\Sigma t)^T \Sigma^{-1} (\Sigma t) - (\Sigma t)^T \Sigma^{-1} (x - \mu) - (x - \mu)^T \Sigma^{-1} (\Sigma t) - 2(t^T \mu + \frac{1}{2} t^T \Sigma t) \\
&= (x - \mu)^T \Sigma^{-1} (x - \mu) + t^T \Sigma t - t^T (x - \mu) - (x - \mu)^T t - 2(t^T \mu + \frac{1}{2} t^T \Sigma t) \\
&= (x - \mu)^T \Sigma^{-1} (x - \mu) - 2t^T (x - \mu) - 2(t^T \mu + \frac{1}{2} t^T \Sigma t) \\
&= (x - \mu)^T \Sigma^{-1} (x - \mu) - 2t^T x
\end{aligned}$$

as desired. □

12.3 Joint cdf from pdf

Example 12.1. Let $f(x, y) = 2e^{-x}e^{-y}$ for $0 < x < y$.

We want to find $F(x, y) = P(X \leq x, Y \leq y)$.

If $x \leq 0$ or $y \leq 0$ then $F(x, y) = 0$ (does not intersect our support set).

When $0 < x \leq y$ (note $P(x = y) = 0$ so it does not matter/influence or cdf), we have

$$\int_0^x \int_x^y f(x, y) dy dx = (1 - e^{-2x}) - 2e^{-y}(1 - e^{-x})$$

For $0 < y < x$, we note that since $x < y$ in our support set we are really calculating $F(y, y)$ so from above

$$F(y, y) = (1 - e^{-2y}) - 2e^{-y}(1 - e^{-y}) = 1 + e^{-2y} - 2e^{-y}$$

If we want to find $F_1(x) = \lim_{y \rightarrow \infty} F(x, y)$.

Note for the region $x \leq 0$, we have $F_1(x) = 0$.

For the region $0 < x < \infty$, we take our $F(x, y)$ and take the limit thus we get $F_1(x) = 1 - e^{-2x}$.

13 October 12, 2018

13.1 Marginal pdf of bivariate normal

Exercise 13.1. $X \sim N(\mu, \sigma^2)$ then $M_X(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}$ for $t \in \mathbb{R}$.

Claim. If $X \sim BVN(\mu, \Sigma)$ then $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$.

Proof. Note that

$$\begin{aligned}
 M_X(t_1, t_2) &= \exp(t^T \mu + \frac{1}{2} t^T \Sigma t) \\
 &= \exp((t_1, t_2)(\mu_1, \mu_2)^T + \frac{1}{2}(t_1, t_2) \begin{bmatrix} \sigma_1^2 & P\sigma_1\sigma_2 \\ P\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} (t_1, t_2)^T) \\
 &= \exp(t_1\mu_1 + t_2\mu_2 + \frac{1}{2}t_1^2\sigma_1^2 + \frac{1}{2}t_2^2\sigma_2^2 + t_1t_2P\sigma_1\sigma_2) \\
 &= \exp(t_1\mu_1 + \frac{1}{2}t_1^2\sigma_1^2) \exp(t_2\mu_2 + \frac{1}{2}t_2^2\sigma_2^2) \exp(t_1t_2P\sigma_1\sigma_2)
 \end{aligned}$$

Thus

$$\begin{aligned}
 M_{X_1}(t_1) &= M_X(t_1, 0) = \exp(t_1\mu_1 + \frac{1}{2}t_1^2\sigma_1^2) \exp(0) \exp(0) \\
 &= \exp(t_1\mu_1 + \frac{1}{2}t_1^2\sigma_1^2)
 \end{aligned}$$

which is the mgf of $N(\mu_1, \sigma_1^2)$. Similarly $M_{X_2}(t_2)$ has the same mgf as $N(\mu_2, \sigma_2^2)$. □

Claim. X_1 and X_2 are independent iff $P = 0$.

Proof. Recall X_1, X_2 are independent iff $M_X(t_1, t_2) = M_{X_1}(t_1)M_{X_2}(t_2)$ for all $t_1, t_2 \in \mathbb{R}$.

Thus the LHS is $M(t_1, t_2) = M_{X_1}(t_1)M_{X_2}(t_2) \exp(t_1t_2P\sigma_1\sigma_2)$ and the RHS is $M_{X_1}(t_1)M_{X_2}(t_2)$ therefore LHS = RHS iff $\exp(t_1t_2P\sigma_1\sigma_2) = 1$ i.e. $t_1t_2P\sigma_1\sigma_2 = 0$ for all $t_1, t_2 \in \mathbb{R}$, thus $P = 0$ since $\sigma_1, \sigma_2 > 0$. □

Claim. If $Y = c^T X \sim N(c^T \mu, c^T \Sigma c)$ where $c = (c_1, c_2)^T \neq (0, 0)^T$.

Proof. Note that $M_Y(t) = E[\exp(tY)] = E[\exp(tc^T X)] = E[\exp((tc)^T X)]$. Let $t^* = tc$ then we have $M_Y(t) = E[\exp((t^*)^T X)] = M(t_1^*, t_2^*) = \exp(t^* \mu + \frac{1}{2}(t^*)^T \Sigma t^*)$.

Thus we have

$$\begin{aligned}
 M_Y(t) &= \exp((tc)^T \mu + \frac{1}{2}(tc)^T \Sigma (tc)) \\
 &= \exp(t(c^T \mu) + \frac{1}{2}t^2(c^T \Sigma c))
 \end{aligned}$$

which is the mgf of $N(c^T \mu, c^T \Sigma c)$ due to the uniqueness theorem of mgf. □

Claim. Let $Y = AX + b$ where $A \in \mathbb{R}^{2 \times 2}$ and $b = (b_1, b_2)^T$.

Then $Y \sim BVN(A\mu + b, A\Sigma A^T)$.

Proof. Exercise (similar to proof above). □

14 October 15, 2018

14.1 Bivariate transformation

Suppose we wanted to transform a bivariate r.v. $(X, Y) \rightarrow (U, V)$ or to U only. We can then find the distribution of (U, V) (or U only) based on (X, Y) .

There are two methods that are analogous to the ones for univariate random variables

Method 1 1-to-1 transformation $(X, Y) \iff (U, V)$

Method 2 cdf technique

Recall in the univariate case with the 1-to-1 technique:

Example 14.1. Let $f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}}$ for $x > \alpha$ ($\alpha, \beta > 0$).

Find the pdf of $Y = \beta \log(\frac{X}{\alpha})$.

Note that $Y = h(X) = \beta \log(\frac{X}{\alpha})$ is a 1-to-1 function as $\alpha, \beta > 0$ and $\log(\cdot)$ is monotonically increasing.

Thus $X = \alpha e^{\frac{Y}{\beta}} = h^{-1}(Y)$. Note that

$$\frac{d}{dy}h^{-1}(y) = \frac{\alpha}{\beta}e^{\frac{y}{\beta}}$$

Thus we have

$$g(y) = f(h^{-1}(y)) \left| \frac{d}{dy}h^{-1}(y) \right| = \frac{\beta\alpha^\beta}{(\alpha e^{\frac{y}{\beta}})^{\beta+1}} \frac{\alpha}{\beta} e^{\frac{y}{\beta}} = e^{-y}$$

Note that the support set of Y is $y > 0$.

Similarly for this univariate case

Example 14.2. Let $X \sim EXP(1)$, $X > 0$. Let $Z = X^2$. Find the pdf of Z .

Method 1 $Z = h(X) = X^2$ is 1-to-1 since X is positive. Thus $X = +\sqrt{Z} = h^{-1}(Z)$ thus $\frac{d}{dz} = \frac{1}{2}z^{-0.5}$ and thus

$$g(z) = e^{-z^{0.5}} \frac{1}{2}z^{-0.5} \quad z > 0$$

Method 2 Note that $G(z) = P(Z \leq z)$, which is 0 if $z \leq 0$.

For $z > 0$ we have

$$\begin{aligned} P(Z \leq z) &= P(X^2 \leq z) = P(X \leq \sqrt{z}) && X \text{ is positive} \\ &= F(\sqrt{z}) \\ &= 1 - e^{-\sqrt{z}} \end{aligned}$$

where $F(x) = 1 - e^{-x}$ if $x > 0$ and 0 if $x \leq 0$.

Thus

$$g(z) = \frac{d}{dz}G(z) = f(\sqrt{z}) \frac{1}{2}z^{-0.5} = e^{-\sqrt{z}} \frac{1}{2}z^{-0.5}$$

where $z > 0$.

What about a bivariate case?

Example 14.3. Recall that with $f(x, y) = ke^{-x}e^{-y}$ for $0 < x < y$ we wanted to find $P(X + Y \geq 1)$, i.e. if $U = X + Y$ we are finding $P(U \geq 1)$. However $(X, Y) \rightarrow U$ is not a 1-to-1 transformation: one pair (X, Y) corresponds to one U but one U does not correspond to a unique (X, Y) , thus we cannot transform our bivariate to a univariate distribution.

Example 14.4. Let $f(x, y) = 3y$ for $0 < x < y < 1$. Find the pdf of $U = XY$ i.e. we want to map $(X, Y) \rightarrow U$.

Note that

$$\begin{aligned} G(u) &= P(U \leq u) \\ &= P(XY \leq u) \\ &= \begin{cases} 0 & \text{if } u \leq 0 \\ (*) & \text{if } 0 < u < 1 \\ 1 & \text{if } u \geq 1 \end{cases} \end{aligned}$$

where the above follows since the support set is $0 < x < y < 1 \Rightarrow 0 < x, y < 1$.

We have (*) as $P(xY \leq u) = P(Y \leq \frac{u}{x})$ for $0 < u < 1$.

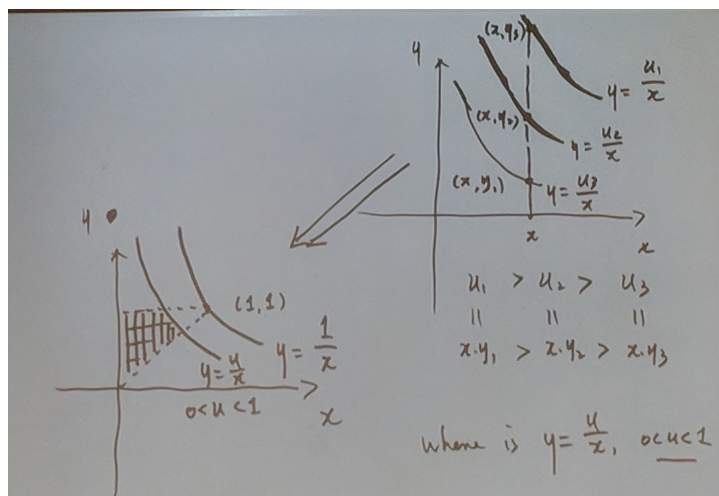


Figure 14.1: Graphs of $y = \frac{u_i}{x}$ for various $0 < u_i < 1$ (top right); graph of $y = \frac{1}{x}$, $y = \frac{u}{x}$ for $0 < u < 1$ and the support triangle region (bottom left). We want to integrate over the shaded area to find $P(Y \leq \frac{u}{x})$.

For a given fixed x , if we look at the $y = \frac{u}{x}$ for $0 < u < 1$, we see that for $P(Y \leq \frac{u}{x})$ we are essentially integrating the area underneath $y = \frac{u}{x}$ that intersects with our support which is the region in the unit square of the first quadrant above $y = x$ since $y > x$.

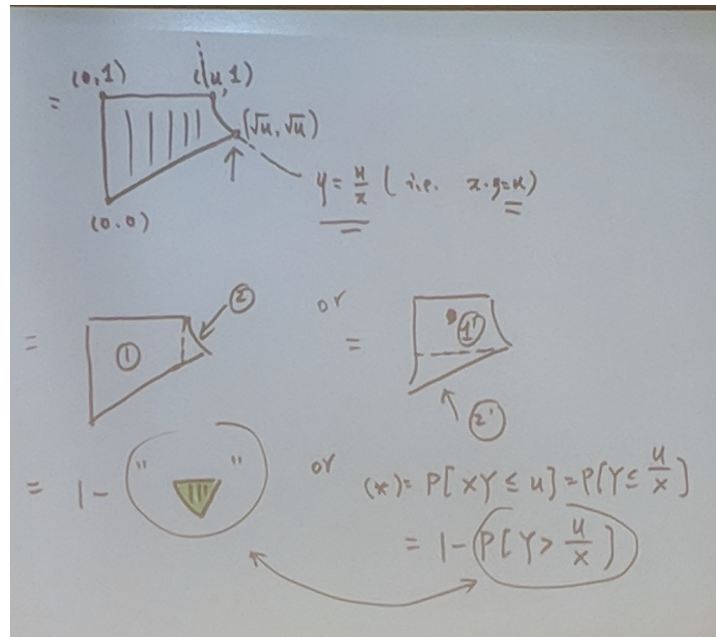


Figure 14.2: Area of integration where we can either integrate over region (1) and (2), (1') and (2'), or integrate over the green triangle and subtract 1 from it.

Note the limits of integration in the image

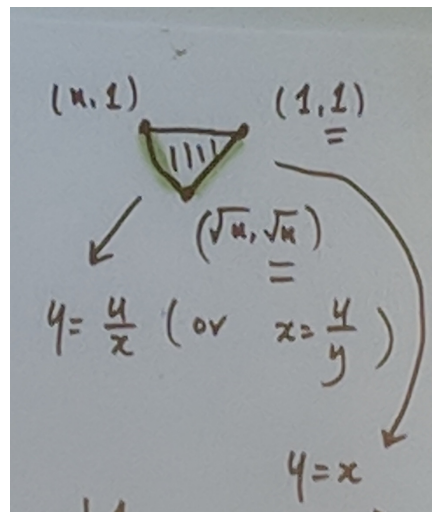


Figure 14.3: Bounds for integrations for the green shaded we are integrating over.

So we'd like to find

$$\begin{aligned}
 P(Y \leq \frac{u}{x}) &= 1 - P(Y > \frac{u}{x}) \\
 &= 1 - \int_{\frac{u}{y}}^1 \left(\int_{\frac{u}{y}}^y f(x, y) dx \right) dy \\
 &= 1 - \int_{\frac{u}{y}}^1 \left(\int_{\frac{u}{y}}^y 3y dx \right) dy \\
 &= 1 - \int_{\frac{u}{y}}^1 3y(y - \frac{u}{y}) dy \\
 &= 1 - \left(y^3 \Big|_{\frac{u}{y}}^1 - 3u(1 - \sqrt{u}) \right) \\
 &= 3u - 2u\sqrt{u} \quad 0 < u < 1
 \end{aligned}$$

Thus we have

$$G(u) = \begin{cases} 0 & \text{if } u \leq 0 \\ 3u - 2u\sqrt{u} & \text{if } 0 < u < 1 \\ 1 & \text{if } u \geq 1 \end{cases}$$

so $g(u) = \frac{d}{du}G(u) = 3 - 3\sqrt{u}$ where $G(u)$ is differentiable for $0 < u < 1$.

$G(u) = 0$ if $u \leq 0$ implies $g(u) = 0$ if $u < 0$, similarly if $u \geq 1$ we have $g(u) = 0$.

What happens if $u = 0$ or $u = 1$? Note that at $u = 0$ $G(u)$ is NOT differentiable. Question: Is $G(u)$ differentiable at $u = 1$?

Regardless, our answer is $g(u) = 3 - 3\sqrt{u}$ for $0 < u < 1$.

Exercise 14.1. Find the pdf of $V = \frac{Y}{X}$.

Example 14.5. Let X_i be iid with common pdf and cdf $f(x)$ and $F(x)$, respectively, $i = 1, \dots, n$.

Find the pdf of $S = \max(X_1, \dots, X_n)$ and $T = \min(X_1, \dots, X_n)$ separately.

Notice that $(x_1, \dots, x_n) \rightarrow S$ (or T) is NOT 1-1.

So

$$\begin{aligned}
 G(s) &= P(S \leq s) = P(\max(X_1, \dots, X_n) \leq s) \\
 &= P(X_1 \leq s, X_2 \leq s, \dots, X_n \leq s) \\
 &= P(X_1 \leq s) \cdot P(X_2 \leq s) \cdot \dots \cdot P(X_n \leq s) && \text{independence} \\
 &= F(s)^n
 \end{aligned}$$

Thus

$$g(s) = \frac{d}{ds}G(s) = nF(s)^{n-1}f(s)$$

where $F(s)$ and $f(s)$ are the cdf and pdf of X evaluated at s , respectively.

Exercise 14.2. Find the pdf of T .

For 1-to-1 bivariate transformation where we have $(X, Y) \iff (U, V)$, we can find the joint pdf of (U, V) based on the joint pdf of (X, Y) .

Recall for $Y = h(X)$ where h is 1-to-1, we have $X = h^{-1}(Y)$ and

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

For $U = h_1(X, Y)$ and $V = h_2(X, Y)$ where h_1, h_2 are 1-to-1, then $X = w_1(U, V)$ and $Y = w_2(U, V)$ where w_1, w_2 are the inverses of h_1, h_2 , respectively.

Thus

$$g(u, v) = f(w_1, w_2) \left| \frac{\partial(w_1, w_2)}{\partial(u, v)} \right|$$

where

$$\frac{\partial(w_1, w_2)}{\partial(u, v)} = \det \begin{bmatrix} \frac{\partial w_1}{\partial u} & \frac{\partial w_1}{\partial v} \\ \frac{\partial w_2}{\partial u} & \frac{\partial w_2}{\partial v} \end{bmatrix} = \frac{\partial w_1}{\partial u} \cdot \frac{\partial w_2}{\partial v} - \frac{\partial w_1}{\partial v} \cdot \frac{\partial w_2}{\partial u}$$

Let R_{XY} and R_{UV} be the support set of (X, Y) of (U, V) , respectively.

Notice that R_{UV} is based on R_{XY} through the bivariate transformation.

Thus the steps for the 1-to-1 technique for bivariate transformations are

1. Verify 1-to-1 transformation
2. Find (w_1, w_2) and $\frac{\partial(w_1, w_2)}{\partial(u, v)}$
3. Find $g(u, v)$
4. Find R_{UV}

15 October 17, 2018

15.1 Verifying 1-to-1 for bivariate functions

The **inverse mapping theorem** states that $U = h_1(X, Y)$ and $V = h_2(X, Y)$ are 1-1 if

1. $\frac{\partial h_1}{\partial x}, \frac{\partial h_1}{\partial y}, \frac{\partial h_2}{\partial x}, \frac{\partial h_2}{\partial y}$ are continuous functions of x and y in R_{XY} .
- 2.

$$\frac{\partial(w_1, w_2)}{\partial(u, v)} = \det \begin{bmatrix} \frac{\partial w_1}{\partial u} & \frac{\partial w_1}{\partial v} \\ \frac{\partial w_2}{\partial u} & \frac{\partial w_2}{\partial v} \end{bmatrix} \neq 0$$

in R_{XY} .

15.2 Examples of bivariate transformations

Example 15.1. Let $X \sim GAM(a, 1)$ independent of $Y \sim GAM(b, 1)$.

Let $U = X + Y = h_1(X, Y)$, $Y = \frac{X}{X+Y} = h_2(X, Y)$.

Find the joint pdf of (U, V) : $g(u, v)$.

Solution. We can do this in 4 parts:

Step 1: Find $f(x, y)$ Note that $f(x, y) = f_1(x)f_2(y)$ by independence so

$$\frac{1}{\Gamma(a)} x^{a-1} e^{-x} \cdot \frac{1}{\Gamma(b)} y^{b-1} e^{-y}$$

Furthermore

$$R_{XY} = R_X \times R_Y = (0, \infty) \times (0, \infty) = \{(x, y) \mid x, y > 0\}$$

Step 2: Verify 1-to-1 Verify 1-to-1 of h_1, h_2 by inverse mapping theorem

1. Note $\frac{\partial h_1}{\partial x} = 1$, $\frac{\partial h_1}{\partial y} = 1$. Also

$$\begin{aligned}\frac{\partial h_2}{\partial x} &= \frac{(x+y) - x}{(x+y)^2} = \frac{y}{(x+y)^2} \\ \frac{\partial h_2}{\partial y} &= \frac{-x}{(x+y)^2}\end{aligned}$$

Note both are continuous on R_{XY} (no discontinuity on R_{XY} since $x+y \neq 0$ and they are quotients of continuous functions which is continuous).

- 2.

$$\begin{aligned}\frac{\partial(h_1, h_2)}{\partial(x, y)} &= \det \begin{bmatrix} \frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y} \\ \frac{\partial h_2}{\partial x} & \frac{\partial h_2}{\partial y} \end{bmatrix} \\ &= \frac{\partial h_1}{\partial x} \cdot \frac{\partial h_2}{\partial y} - \frac{\partial h_1}{\partial y} \cdot \frac{\partial h_2}{\partial x} \\ &= 1 \cdot \frac{-x}{(x+y)^2} - 1 \cdot \frac{y}{(x+y)^2} \\ &= \frac{-1}{x+y} \\ &\neq 0 \quad (x, y) \in R_{XY}\end{aligned}$$

Therefore our functions are indeed 1-to-1 by the inverse mapping theorem.

Step 3: Find inverse $g(u, v)$ We find our inverse transformations and $\frac{\partial(w_1, w_2)}{\partial(u, v)}$. Note that we can see

$$\begin{aligned}X &= UV = w_1(U, V) \\ Y &= U - UV = U(1 - V) = w_2(U, V)\end{aligned}$$

We also have

$$\begin{aligned}\frac{\partial(w_1, w_2)}{\partial(u, v)} &= \left(\frac{\partial w_1}{\partial u}\right)\left(\frac{\partial w_2}{\partial v}\right) - \left(\frac{\partial w_1}{\partial v}\right)\left(\frac{\partial w_2}{\partial u}\right) \\ &= (v)(-u) - (1-v)(u) \\ &= -u\end{aligned}$$

So we have (where $f(x, y) = \frac{1}{\Gamma(b)\Gamma(a)}x^{a-1}e^{-x}y^{b-1}e^{-y}$)

$$\begin{aligned}g(u, v) &= f(w_1, w_2) \left| \frac{\partial(w_1, w_2)}{\partial(u, v)} \right| \\ &= \frac{1}{\Gamma(b)\Gamma(a)} (uv)^{a-1} e^{-uv} (u(1-v))^{b-1} e^{-u(1-v)} \cdot |-u| \\ &= \frac{1}{\Gamma(b)\Gamma(a)} u^{a+b-1} e^{-u} v^{a-1} (1-v)^{b-1}\end{aligned}$$

Remark 15.1. We can factorize $g(u, v) = f_1(u)f_2(v)$ into a function of u and a function of v .

Step 4: Find R_{UV} support We derive this from R_{XY} : note that $R_{XY} = \{(x, y) \mid x, y > 0\}$, where $X = UV$ and $Y = U - UV$, thus we have

$$R_{UV} = \{(u, v) \mid w_1(u, v) = uv > 0, w_2(u, v) = u - uv > 0\}$$

Since $uv > 0$, then $u, v > 0$.

Secondly since $u - uv > 0$, then $u > uv > 0$ so $u, v > 0$ and $v < 1$.

Thus we have $R_{UV} = \{(u, v) \mid u > 0, 0 < v < 1\}$.

That is $R_{UV} = (0, \infty) \times (0, 1)$ is rectangular, so U, V are independent by the remark above and the factorization theorem.

Optional step: Marginal pdfs We claim $U \sim \text{GAM}(a + b, 1)$ and $V \sim \text{BETA}(a, b)$. Note that

$$g_1(u) = \int_0^1 g(u, v) dv = \frac{1}{\Gamma(a + b)} u^{a+b-1} e^{-u}$$

which is the pdf of $\text{GAM}(a + b, 1)$.

Similarly

$$g_2(v) = \int_0^\infty g(u, v) du = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} v^{a-1} (1 - v)^{b-1}$$

which is the pdf of $\text{BETA}(a, b)$.

15.3 Computing determinant from the inverse's determinant

Note that in the previous example, we can compute $\frac{\partial(w_1, w_2)}{\partial(u, v)}$ indirectly by

$$\frac{\partial(w_1, w_2)}{\partial(u, v)} = \left[\frac{\partial(h_1, h_2)}{\partial(x, y)} \right]^{-1} \Big|_{x=w_1(u, v), y=w_2(u, v)}$$

where we need to substitute x and y for $w_1(u, v)$ and $w_2(u, v)$ after computing the inverse of the determinant. Recall from our previous example that we had

$$\frac{\partial(h_1, h_2)}{\partial(x, y)} = \frac{-1}{x + y}$$

thus we have

$$\frac{\partial(w_1, w_2)}{\partial(u, v)} = \left(\frac{-1}{x + y} \right)^{-1} = -(x + y) = -(uv + u - uv) = -u$$

which agrees with our previous result.

15.4 Bivariate transformations of non 1-to-1 functions

Example 15.2. For $f(x, y) = 3y$ where $0 < x < y < 1$, let $U = XY$. Suppose we wanted to find the pdf of U . Note that $(X, Y) \rightarrow U$ is not 1-to-1: we have multiple $f(x, y)$ for the same u . For example if we had $u = 16$, we can either $(x, y) = (2, 8)$ or $(x, y) = (4, 4)$ which maps to different $f(x, y)$ values.

We can include some random variable V to ensure $(X, Y) \leftrightarrow (U, V)$ is 1-to-1, then we can compute the marginal pdf of U via $g_1(u) = \int g(u, v) dv$.

What V do we choose? We claim V is not unique. Let $V = X$, so $U = h_1(X, Y) = XY$ and $V = h_2(X, Y) = X$.

We note that for $u = 16, v = 2$, we only have one $(x, y) = (2, 8)$ that maps to one unique $f(x, y) = 24$ value. Similarly $V = Y$ works as well.