BIOS 660/BIOS 672 (3 Credits):
Probability and Statistical Inference I

Jianwen Cai

https://sakai.unc.edu/portal/site/bios660-bios672-3-credits

Notes 13

Multiple Possible Outcomes of a Trial

Suppose: Result of drug trial is Failure, Partial Success and Success. Let

$$P(F) = p_1, \quad P(PS) = p_2, \quad P(S) = p_3, \quad (p_1 + p_2 + p_3) = 1$$

Suppose in a sample of size n

$$\begin{aligned} s_1 &= \text{Number of Failures} \\ s_2 &= \text{Number of Partial Successes} \\ s_3 &= \text{Number of Successes} \\ where \quad s_1 + s_2 + s_3 = n, & \\ P(s_1, s_2, s_3) &= \frac{n!}{s_1!s_2!s_3!} p_1^{s_1} p_2^{s_2} p_3^{s_3} \end{aligned}$$

These are *multinomial probabilities*.

continued

The probability of any ordered arrangement resulting in s_1"F", s_2 "PS" and s_3 "S" is

$$p_1^{s_1} p_2^{s_2} p_3^{s_3}$$

However there are $\frac{n!}{s_1!s_2!s_3!}$ such ordered arrangements. Therefore

$$P(s_1, s_2, s_3) = \frac{n!}{s_1!s_2!s_3!} p_1^{s_1} p_2^{s_2} p_3^{s_3}$$

Multinomial distribution

The generalization to k classes gives us the *Multinomial Distribution*

$$p(s_1, s_2, \ldots, s_k) = \frac{n!}{s_1!s_2!\ldots, s_k!} \; p_1^{s_1} p_2^{s_2} \ldots p_k^{s_k}$$

where $\sum_{i=1}^{k} s_i = n$ and $\sum_{i=1}^{k} p_i = 1$.

The Poisson distribution again

sample space: $\{0, 1, 2...\}$
pmf:

$$P(s) = \begin{cases} e^{-\lambda}\lambda^s/s! & s = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

cdf:

$$F(y) = \sum_{s=0}^{y} e^{-\lambda}\lambda^s/s!$$

expectation:

$$\mathsf{E}(Y) = \lambda$$

Variance:

$$\mathsf{Var}(Y) = \lambda$$

continued

Note: we can write

$$\frac{P(X=i+1)}{P(X=i)} = \frac{\lambda}{i+1}$$

Poisson: Examples and generalizations

Last time, we had an example concerning pulmonary embolism among young women. The rate is 4 per million, and we had looked at the pdf for the number of cases in a city with 1,000,000 women. But, suppose we are interested in a city that only has 100,000 women. How does the probability distribution change?

$$p = \frac{4}{1,000,000}, \quad n = 100,000, \quad np = \frac{4}{10} = .4 = \lambda$$

$P(s) = e^{-.4}(.4)^s/s!$

s	$p(s)$
0	$.67$
1	$.27$
2	$.05$
3	$.007$
4	$.0007$

continued

It is often useful to write

$P(s) = e^{-\lambda n}(\lambda n)^s/s!$ λ = rate/unit population
n = size of population

where λn is the expected number of events.
Note: "population" can be in units of 10, 100, 1000, etc. λ is the mean number of events per unit population.

Example: Failure of Equipment

A computer has a failure rate of 1 failure per 1,000 hours of use. How many failures would be expected in 500 hours of use?

What is the probability distribution of the number of failures?

$\underline{s}$	$P(s) = e^{-.5}(.5)^s/s!$
0	.61
1	.30
2	.08
3	.01

More generally, the distribution may be written

$$P(s) = e^{-\lambda t}(\lambda t)^s/s!$$

where λt = number of failures (events) in t units of time.

Now suppose there are n computers, each being observed for t time units and each having a failure rate of λ, what is $P(s)$?

Continued

$P(s) = e^{-\lambda nt}(\lambda nt)^s/s!$

λ	$=$	rate per unit time per unit individual
n	$=$	number of units (individuals)
t	$=$	time frame
λnt	$=$	Expected number of events for n units and time t

Example: Leukemia

Suppose the incidence rate for childhood leukemia is 3.9 per 100,000 per year for children less than 4 years old. In a population of 5,000 children observed for 10 years, what would the probability distribution be for number of cases?

$$\lambda nt = \frac{3.9}{100,000} \times 5,000 \times 10 = 1.95$$

$$P(s) = e^{-1.95}(1.95)^s/s!$$

s	$P(s)$
0	.14
1	.28
2	.27
3	.18
4	.08
5	.03
6	.01
7	.003

Example: Safety Testing of Vaccine

Suppose a vaccine contains m live virus per cm^3. Suppose a sample of v cm^3 of vaccine is tested. The expected number of virus in v cm^3 is thus equal to mv. What is probability that vaccine tested will be free of a virus?

$$P(s) = e^{-mv}(mv)^s/s!$$

$$P(0) = e^{-mv}$$

e.g. Suppose $m = .005$ and $v = 600cc$. Then, $mv = 3$ and $P(0) = e^{-3} = .05$

Example: Leukemia in Woburn, MA

During a 19 year period 15 leukemias were observed. Is this an unusual event?

Age	Population (n)	Rate per $10^5/yr(\lambda)$	Expected Number
$0-4$	2120	6.27	.133
$5-9$	2191	3.09	.068
$10-14$	2969	2.04	.061
$15-19$	3592	2.19	.079
	$10,872$		.341

During a 19 year period, we expect (19)(.341) = 6.5

Therefore, $P(s) = e^{-6.5}(6.5)^s/s!$ and $\sum_{s=15}^{\infty} e^{-6.5}(6.5)^s/s! = .007$

Probability of observing 15 or more leukemias = .007.

continued

Note: For this example, we are glossing over the fact here that we are really combining several different populations that have different Poisson rates. We will see that if X_1 and X_2 are two independent Poisson random variables with parameters λ_1 and λ_2, then their sum is Poisson with parameter $\lambda_1 + \lambda_2$.

More Negative Binomial

Negative Binomial

In the context of iid Bernoulli trials, define a random variable corresponding to the number of trials required to have s successes. We say $Y \sim$ *Negbin*(s, p):
sample space: $\{s, (s+1), ...\}$
pmf: for $y = s, s+1, s+2, \ldots,$

$$f(y) = \binom{y-1}{s-1} p^s q^{y-s}$$

Why? If y is number of trials, then first $(y-1)$ trials resulted in $(s-1)$ successes and the last trial is success.
cdf: no closed form.
expectation: $\mathrm{E}(Y) = s/p$
Variance: $\mathrm{Var}(Y) = s(1-p)/p^2$
Recall that the negative binomial is the sum of s independent geometrics with parameter p such that these follow immediately.

Example

The Red Sox and the Atlanta Braves are playing in the world series. The winning team is the first one to win 4 games. Suppose each game is independent of the others, and that the Red Sox win a game with probability p. What is the probability that the Red Sox win?

$$\begin{aligned} P(\text{Red Sox wins}) &= \binom{3}{3} p^4 + \\ &\binom{4}{3} p^4(1-p) + \\ &\binom{5}{3} p^4(1-p)^2 + \\ &\binom{6}{3} p^4(1-p)^3 \end{aligned}$$

Example (continued)

What is the probability that the series goes to 7 games?
If the Red Sox wins, then we need to have three Braves wins and three Red Sox wins (in any order) followed by a Red Sox win:

$$\binom{6}{3} p^4(1-p)^3.$$

Similarly, the probability that the Braves wins in 7 games is

$$\binom{6}{3} p^3(1-p)^4.$$

Hence the total probability is

$$\binom{6}{3} [p^4(1-p)^3 + p^3(1-p)^4].$$

Comments

- One of the most important things to get out of this class is to understand the different distributions and when/where you would choose to use them.
- You need to be very familiar with all of these distributions as well as simple transformations of these distributions e.g. what happens if you scale an exponential? what about a gamma?
- You will also need to be very familiar with what happens if there are multiple random variables (next class) and need to understand what happens if they are combined (transformed), e.g. what happens if you add Poissons, normals, exponentials? what happens if you take the ratio of a binomial to a Poisson?

5

Exponential Families

Exponential Families

A family of pdfs or pmfs with vector parameter $\boldsymbol{\theta}$ is called an *exponential family* if it can be expressed as

$$f(x|\boldsymbol{\theta}) = h(x)\, c(\boldsymbol{\theta})\, \exp\left(\sum_{j=1}^{k} w_j(\boldsymbol{\theta})\, t_j(x)\right), \qquad x \in S \subset \mathbb{R}$$

where

- S is not defined in terms of $\boldsymbol{\theta}$ (i.e., **the support of the distribution does not depend on the unknown parameter**)
- $h(x), c(\boldsymbol{\theta}) \geq 0$ and the functions are just functions of the parameters specified; i.e. h is free of $\boldsymbol{\theta}$, $c(\boldsymbol{\theta})$ is free of x, etc...

Example: Binomial

Let $X \sim Binom(n,p)$, $0 < p < 1$.

$$
\begin{aligned}
f(x|p) &= \binom{n}{x} p^x \, (1-p)^{n-x} = \binom{n}{x} (1-p)^n \, \left[\frac{p}{1-p}\right]^x \\
&= \binom{n}{x} (1-p)^n \, \exp\left[\log\left(\frac{p}{1-p}\right)\, x\right]
\end{aligned}
$$

Thus,

$$
\begin{aligned}
h(x) &= \binom{n}{x}, \quad x = 0, \ldots, n \qquad w_1(p) = \log\left(\frac{p}{1-p}\right) \\
c(p) &= (1-p)^n, \quad 0 < p < 1 \qquad t_1(x) = x
\end{aligned}
$$

Note that this works when p is considered the parameter, while n is fixed. If n is not fixed, then the support depends on the unknown parameter. Also, p cannot be 0 or 1.

Example: Gaussian

Let $X \sim N(\mu, \sigma^2)$.

$$
\begin{aligned}
f_X(x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{\mu^2}{2\sigma^2}\right) \exp\left(- \frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\right)
\end{aligned}
$$

Thus

$$
\begin{aligned}
h(x) &= \frac{1}{\sqrt{2\pi}} \qquad c(\mu,\sigma) = \frac{1}{\sigma} \exp\left(- \frac{\mu^2}{2\sigma^2}\right) \\
w_1(\mu,\sigma) &= -\frac{1}{2\sigma^2} \qquad w_2(\mu,\sigma) = \frac{\mu}{\sigma^2} \\
t_1(x) &= x^2 \qquad t_2(x) = x
\end{aligned}
$$

The parameter space is $(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$.

Theorem C-B 3.4.2

If X is a rv from the exponential family, then

$$\mathsf{E}\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j}\, t_i(X)\right) = -\frac{\partial}{\partial \theta_j}\, \log c(\boldsymbol{\theta})$$

$$\begin{aligned}\mathsf{Var}\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j}\, t_i(X)\right) &= -\frac{\partial^2}{\partial \theta_j^2} \log c(\boldsymbol{\theta}) \\ &- \mathsf{E}\left(\sum_{i=1}^{k} \frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial \theta_j^2}\, t_i(X)\right)\end{aligned}$$

Proof: Homework

Example: Binomial

Recall:

$$w_1(p) = \log \frac{p}{1-p}, \quad c(p) = (1-p)^n$$

Relevant derivatives:

$$\frac{\partial}{\partial p} w_1(p) = \frac{\partial}{\partial p} \log \frac{p}{1-p} = \frac{1}{p(1-p)}$$

$$\frac{\partial}{\partial p} \log c(p) = \frac{\partial}{\partial p} n \log(1-p) = \frac{-n}{1-p}$$

So

$$\mathsf{E}\left[\frac{1}{p(1-p)} X\right] = \frac{n}{1-p} \quad \Rightarrow \quad \mathsf{E}(X) = np$$

Indicator function

Definition 3.4.5. The *indicator function* of a set A, most often denoted by $I_A(x)$, is the function

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases}$$

Also denoted as $I(x \in A)$, $1_A(x)$, or $1(x \in A)$.

Note on exponential family:

- The set of x values for which $f(x|\boldsymbol{\theta}) > 0$ cannot depend on $\boldsymbol{\theta}$ in an exponential family.
- The entire definition of the pdf or pmf must be incorporated into the form for the exponential family.
- Incorporate the range of x into the expression for $f(x|\boldsymbol{\theta})$ through the use of an indicator function.

Example. Normal pdf:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) I_{(-\infty,\infty)}(x).$$

More examples

Other exponential families are: Poisson, negative binomial, gamma, beta.

Some densities that are not exponential families: t, F.

Uniform: $X \sim U(0, \theta)$

$$f_X(x) = \theta^{-1} I(0 < x < \theta)$$

Truncated exponential:

$$f_X(x) = \theta^{-1} \exp(1 - x/\theta)\, I(\theta, \infty)$$

What about $X \sim U(0, 1)$?

Natural parameters

An exponential family can be reparametrized as

$$f(x|\boldsymbol{\eta}) = h(x)\, c^*(\boldsymbol{\eta})\, \exp\left(\sum_{j=1}^{k} \eta_j\, t_j(x)\right), \qquad x \in S \subset \mathbb{R}$$

where $\boldsymbol{\eta}$ is called the natural parameter vector.

This parametrization is often more useful. We have the following property:

$$\begin{aligned} \mathsf{E}\left[t_j(X)\right] &= -\frac{\partial}{\partial \eta_j}\, \log c^*(\boldsymbol{\eta}) \\ \mathsf{Var}\left[t_j(X)\right] &= -\frac{\partial^2}{\partial \eta_j^2} \log c^*(\boldsymbol{\eta}) \end{aligned}$$

Proof: Homework

Example: Gaussian

Let $X \sim N(\mu, \sigma^2)$. Recall:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\right)$$

and

$$w_1(\mu, \sigma) = -\frac{1}{2\sigma^2}, \;\; w_2(\mu, \sigma) = \frac{\mu}{\sigma^2}$$

Define

$$\eta_1 = -\frac{1}{2\sigma^2} < 0, \qquad \eta_2 = \frac{\mu}{\sigma^2} \in \mathbb{R}$$

then

$$\sigma^2 = -\frac{1}{2\eta_1} > 0, \qquad \mu = -\frac{\eta_2}{2\eta_1} \in \mathbb{R}$$

The parameter space is now $(\eta_1, \eta_2) \in (-\infty, 0) \times \mathbb{R}$.

Curved exponential families

Let d be the dimension of the parameter space of the exponential family with k terms. The exp. family is called

full	if	$d = k$
curved	if	$d < k$

Example: $X \sim N(\mu, \sigma^2)$. Suppose $\sigma^2 = \mu^2$, i.e. the coefficient of variation is constant equal to 1. The parameter space $(\mu, \sigma^2) = (\mu, \mu^2)$ is now a parabola . For the natural parameters:

$$\eta_1 = -\frac{1}{2\mu^2}, \quad \eta_2 = \frac{1}{\mu} \qquad \Rightarrow \qquad \eta_1 = -\frac{\eta_2^2}{2}$$

Example: Let $X_1, \ldots, X_n$ be an iid sample from $Po(\lambda)$. Let $\bar{X} = \sum_{i=1}^n X_i/n$. Then for large n (by the Central Limit Theorem (you will learn this in bios 661/673)),

$$\frac{\bar{X} - \lambda}{\sqrt{\lambda}} \Rightarrow N(0,1) \qquad \text{so} \qquad X \dot{\sim} N(\lambda, \lambda)$$

This is a curved exp. family with parameter space $(\mu, \sigma^2) = (\mu, \mu)$.

Chebychev Inequality

Let X be a random variable and let $g(x)$ be a non-negative function. Then for any $r > 0$,

$$P[g(X) \geq r] \leq \frac{\mathsf{E}g(X)}{r}$$

Proof:

$$\begin{aligned} \mathsf{E}g(X) &= \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx \\ &\geq \int_{\{x:g(x)\geq r\}} g(x)\, f_X(x)\, dx \\ &\geq r \int_{\{x:g(x)\geq r\}} f_X(x)\, dx \\ &= r\, P\{g(X) \geq r\} \end{aligned}$$

Application

Let

$$g(x) = \frac{(x-\mu)^2}{\sigma^2}$$

where $\mu = \mathsf{E}(X)$, $\sigma^2 = \mathsf{Var}(X)$.

Let $r = t^2$, then

$$P\left[\frac{(x-\mu)^2}{\sigma^2} \geq t^2\right] \leq \frac{1}{t^2}\mathsf{E}\left[\frac{(x-\mu)^2}{\sigma^2}\right]$$
$$P[|X-\mu| \geq t\sigma] \leq \frac{1}{t^2}$$

The probability that X is more than $t\sigma$ away from μ cannot be more than $1/t^2$, no matter what the distribution of X. E.g. $t = 2$.

Normal tail bound

Let $Z \sim N(0,1)$:

$$P[Z \geq t] = \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-x^2/2}\,dx$$
$$\leq \frac{1}{\sqrt{2\pi}}\int_t^\infty \frac{x}{t}e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\frac{e^{-t^2/2}}{t}$$

and so

$$P[|Z| \geq t] = 2P[Z \geq t] \leq \sqrt{\frac{2}{\pi}}\frac{e^{-t^2/2}}{t}$$

For $t = 2$, the bound is 0.054.

Multiple random measurements

Multiple endpoints in health studies:

- Cancer: survival, quality of life, toxicity
- Reproductive health: time to pregnancy, birth defects
- AIDS: time from infection to AIDS, time from AIDS to death
- Carcinogenicity studies: time to tumor, time to death
- Health care: cost, hospital duration of stay

Multiple time points/spatial locations:

- Environmental monitoring: daily temperature, humidity, CO_2/ozone concentration, geographically located monitoring stations.
- Finance: daily stock prices, portfolios.

Massively multivariate:

- Genomics and other omics: thousands of gene expression values, SNPs, protein concentrations, metabolite concentrations.
- Imaging: thousands of voxels (volume pixels) measuring brain activity (fMRI), tumor metabolism (PET); satellite remote sensing.

Random Vectors

Suppose we start with a probability space $(\Omega, \mathcal{A}, P)$.

Defintion: An n-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a function from a sample space Ω into $\mathbb{R}^n$.

- Each coordinate X_i is a random variable.
- The random vector is associated with a probability space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), F)$.
- For every Borel set B,
$$P\{\boldsymbol{X} \in B\} = P\{\boldsymbol{X}^{-1}(B)\}$$
  where
$$\boldsymbol{X}^{-1}(B) = \{\omega : \boldsymbol{X}(\omega) \in B\}$$

Example: Bivariate

A fair coin is flipped 3 times. Define the random vector (X, Y) where X represents the number of heads on the last toss and Y the total number of heads. Then, the probabilities of various outcomes are given in the following table:

Outcome	(x, y)	P(outcome)
(H,H,H)	(1,3)	1/8
(H,H,T)	(0,2)	1/8
(H,T,H)	(1,2)	1/8
(H,T,T)	(0,1)	1/8
(T,H,H)	(1,2)	1/8
(T,H,T)	(0,1)	1/8
(T,T,H)	(1,1)	1/8
(T,T,T)	(0,0)	1/8

Discrete Bivariate RVs

Two random variables X and Y are said to be jointly *discrete* if there is an associated *joint probability mass function*,

$$f_{X,Y}(x, y) = P\{X = x,\ Y = y\}$$

which sums to 1 over a finite or possibly countable combinations of x and y for which $f_{X,Y}(x, y) > 0$, i.e.,

$$\sum_{x,y} f_{X,Y}(x, y) = 1$$

From this, one can also obtain the marginal pmfs of X and Y as follows:

$$f_X(x) = P(X = x) = \sum_y f_{X,Y}(x, y)$$

$$f_Y(y) = P(Y = y) = \sum_x f_{X,Y}(x, y)$$

Example

Back to the fair coin example again. From the definition we can construct the joint pdf of X and Y:

		Y			
		0	1	2	3
X	0	1/8	1/4	1/8	0
	1	0	1/8	1/4	1/8

The marginal distributions of X and Y are also easy to find.

Note: Marginals do not determine joint pmf.

Bivariate cdfs

Regardless of whether they are discrete or continuous or some combination of the two, we can always define the *joint cumulative distribution function*.

For $n = 2$, the *bivariate cumulative distribution function* is

$$F_{X,Y}(x, y) = P\{X \le x, Y \le y\}$$

Properties:

- $F_{X,Y}(x, y) \ge 0$
- $F_{X,Y}(\infty, \infty) = 1$
- $F_{X,Y}(-\infty, y) = F(x, -\infty) = 0$
- $F_{X,Y}(-\infty, -\infty) = 0$
- F is non-decreasing and right-continuous in each variable separately

Marginal distributions

From $F_{X,Y}$, we can derive the univariate distribution functions for X and Y. These are generally called *marginal distributions*.

$$F_X(x) = P\{X \leq x\} = P\{X \leq x, Y < \infty\} = F_{X,Y}(x, \infty)$$
$$F_Y(y) = P\{Y \leq y\} = P\{X < \infty, Y \leq y\} = F_{X,Y}(\infty, y)$$

Note: Although we can obtain $F_X(x)$ and $F_Y(y)$ from the joint *cdf*, we cannot do the reverse.

Joint probabilities

All joint probability statements about X and Y can be answered in terms of their joint *cdf*. For example,

$$P(X > x, Y > y) = 1 - F_X(x) - F_Y(y) + F_{X,Y}(x, y)$$

More generally,

$$P(x_1 < X \leq x_2,\ y_1 < Y \leq y_2) =$$
$$F_{X,Y}(x_2, y_2) + F_{X,Y}(x_1, y_1) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1)$$

Comments:

Continuous Bivariate Random Variables

Continuous Bivariate RVs

The random variables X and Y are said to be *jointly (absolutely) continuous* if there exists a function $f_{X,Y}(x, y)$, such that for any Borel set B of 2-tuples in $\mathbb{R}^2$,

$$P\{(X, Y) \in B\} = \int\int_{(x,y)\in B} f_{X,Y}(x, y)dxdy$$

The function $f_{X,Y}(x, y)$ is called the *joint probability density function* for X and Y.

It follows in this case that

$$F_{X,Y}(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(s, t)dsdt,$$

$$f_{X,Y}(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

Properties of the bivariate pdf

- $f_{X,Y}(x, y) \geq 0$
- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)dxdy = 1$
- $f_{X,Y}(x, y)$ is **not a probability**, but can be thought of as a relative probability of (X, Y) falling into a small rectangle located at (x, y):

$$P\{x < X \leq x + dx,\;\; y < Y \leq y + dy\} \approx f(x, y)dxdy$$

- The *marginal probability density functions* for X and Y can be obtained as

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$$

Example 1

$$F_{XY}(x,y) = xy \qquad\qquad 0 < x \leq 1, \quad 0 < y \leq 1$$

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x\, \partial y} = 1$$

$$f_X(x) = \int_0^1 dy = 1$$

$$f_Y(y) = \int_0^1 dx = 1$$

Example 2

$$F_{XY}(x,y) = x - x\ \log(\tfrac{x}{y}) \qquad 0 < x \leq y \leq 1$$

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}}{\partial x \partial y} = \frac{\partial}{\partial x}[-x(\tfrac{y}{x})(-\tfrac{x}{y^2})] = \frac{\partial}{\partial x}\frac{x}{y} = \frac{1}{y}$$

$$f_X(x) = \int_x^1 \frac{dy}{y} = -\log(x)$$
$$f_Y(y) = \int_0^y \frac{dx}{y} = 1$$

Note: Once we have $f_X(y)$ and $f_Y(y)$, we can obtain $F_X(x)$ and $F_Y(y)$ directly.

Double check:

$$\begin{aligned} F_X(x) &= F_{X,Y}(x,1) = x - x\log(x); \\ & \frac{d}{dx}[x - x\log(x)] = -\log(x). \\ F_Y(y) &= F_{X,Y}(y,y) = y; \\ & \frac{d}{dy}y = 1. \end{aligned}$$

Conditional Distributions - Discrete

Recall if A and B are two events, the probability of A conditional on B is:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{or} \quad \frac{P(AB)}{P(B)}$$

Defining the events $A = \{Y = y\}$ and $B = \{X = x\}$, it follows that

$$\begin{aligned} P\{Y = y|X = x\} &= \frac{P(X = x,\ Y = y)}{P(X = x)} \\ &= \frac{f_{X,Y}(x, y)}{f_X(x)} \\ &= f_{Y|X}(y|x) \end{aligned}$$

This is called the **conditional probability mass function** of Y given X.

Example: Discrete

A fair coin is flipped 3 times. Define the random vector (X, Y) where X represents the number of heads on the last toss and Y the total number of heads. From the joint pmf of X and Y we can derive all the conditional pmfs: $f_{Y|X}(y|x) = f_{X,Y}(x, y)/f_X(x)$ and $f_{X|Y}(x|y) = f_{X,Y}(x, y)/f_Y(y)$.

Examples: $f_{Y|X}(0|0) = f_{X,Y}(0, 0)/f_X(0) = \frac{1/8}{1/2} = 1/4$;
$f_{X|Y}(1|2) = f_{X,Y}(1, 2)/f_Y(2) = \frac{1/4}{3/8} = 2/3$.

		Y				
		0	1	2	3	Sum
X	0	1/8	1/4	1/8	0	1/2
	1	0	1/8	1/4	1/8	1/2
	Sum	1/8	3/8	3/8	1/8	1

Conditional Distributions - Continuous

If $F(x,y)$ is absolutely continuous, we define the conditional density of Y given X as:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \text{ if } f_Y(y) > 0$$

Because Y is continuous, we cannot directly evaluate this probability, since the denominator will be zero. Instead, think of small dx, dy:

$$\begin{aligned} & Pr(x \le X < x+dx \,|\, y \le Y < y+dy) \\ = \; & \frac{Pr(x \le X < x+dx, y \le Y < y+dy)}{Pr(y \le Y < y+dy)} \\ \approx \; & \frac{f(x,y)dxdy}{f_Y(y)dy} \\ = \; & f_{X|Y}(x|y)dx \end{aligned}$$

Show that it satisfies the conditions for a density.

Example 1

$$\begin{aligned} & F_{XY}(x,y) = xy && 0 < x \le 1, \quad 0 < y \le 1 \\ & f_{XY}(x,y) = 1 && 0 < x < 1, \quad 0 < y < 1 \\ & f_X(x) = 1 && 0 < x < 1 \\ & f_Y(y) = 1 && 0 < y < 1 \\ & f_{X|Y}(x|y) = \tfrac{f_{XY}(x,y)}{f_Y(y)} = 1 && 0 < x < 1 \quad (0 < y < 1) \\ & f_{Y|X}(y|x) = \tfrac{f_{XY}(x,y)}{f_X(x)} = 1 && 0 < y < 1 \quad (0 < x < 1) \end{aligned}$$

In this particular case, we get that the conditional densities are the same as the marginals. This means X and Y are independent.

Example 2

$$F_{XY}(x,y) = x - x\ \log\frac{x}{y} \qquad 0 < x \le y \le 1$$

$$f_{XY}(x,y) = 1/y \qquad 0 < x \le y \le 1$$

$$f_X(x) = -\log x \qquad 0 < x \le 1$$

$$f_Y(y) = 1 \qquad 0 < y \le 1$$

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = 1/y \qquad 0 < x \le y \qquad (0 < y \le 1)$$

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = -\frac{1}{y\log x} \qquad x \le y \le 1 \qquad (0 < x \le 1)$$

- Y is marginally uniform, but not conditionally
- X is conditionally uniform, but not marginally