Lecture 3

Joint distributions

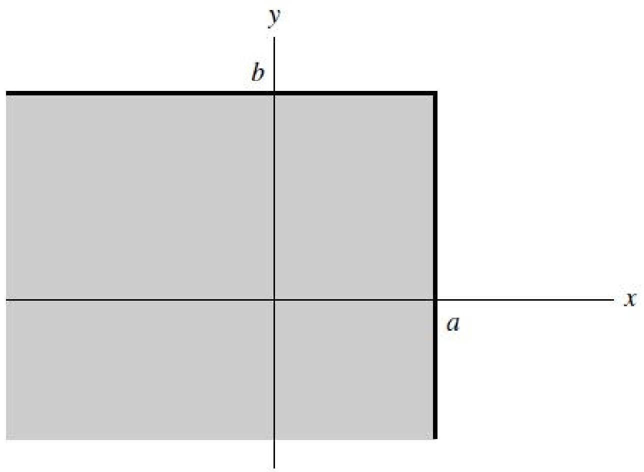
Joint probability structure of two or more random variables defined on the same sample space.

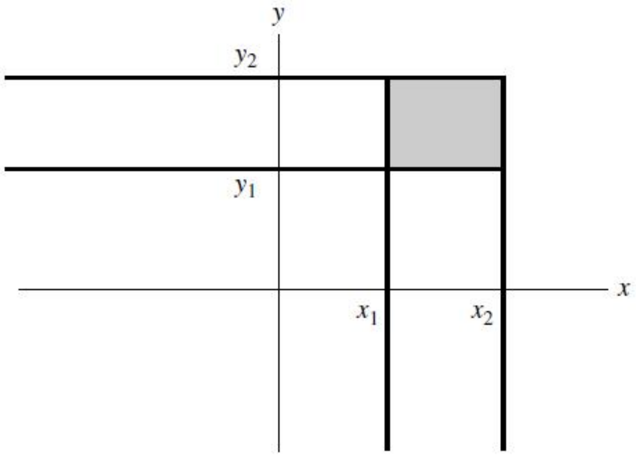
- Turbulent atmosphere/molecular cloud/quasar outflow: in a realistic 3-D model, the joint probability distribution of *x, y, z* components of wind velocity can be measured or simulated.
- Setting fish harvesting policies: a model for the joint distribution of age and length in a population of fish can be used to estimate the age distribution from the length distribution.

The joint behavior of two random variables, *X* and *Y*, is determined by cdf

$$F(x, y) = P(X \leq x, Y \leq y)$$

-- probability that point (*X,Y*) belongs to a semi-infinite rectangle in the plane.





Probability that point (X, Y) belongs to a given rectangle is?

$$P(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$$

More complicated situations, more intersections and unions of rectangles.

In general, if $X_1, \ldots, X_n$ are jointly distributed random variables, their **joint cdf** is

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

Random variables: discrete

X, Y are discrete random variables defined on the same sample space, taking on values $x_1, \ldots, x_n$, $y_1, \ldots, y_n$, their joint frequency/probability mass function is

$$p(x_i, y_j) = P(X = x_i, Y = y_j)$$

- *Examples.*

  Toss a coin 3 times. X=# of heads on the first toss, Y=total # of heads.

$$\Omega = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$$

$$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \Delta \quad \Delta \quad \Delta \quad \Delta$$

Joint frequency function:

x \ y	0	1	2	3
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

Frequency function of Y?

$$p_Y(0) = P(Y = 0) = \frac{1}{8} + 0$$

$$p_Y(1) = P(Y = 1) = \frac{3}{8}$$

Similar for X: $p_X(x) = \sum_i p(x, y_i)$

Marginal frequency function

$X_1, \ldots, X_n$ are defined on the same sample space, their joint frequency function is

$$p(x_1, \ldots, x_m) = P(X_1 = x_1, \ldots, X_m = x_m)$$

Marginal frequency function of X_1 is $$p_{X_1}(x_1) = \sum_{x_2 \cdots x_m} p(x_1, x_2, \ldots, x_m)$$

2-d marginal frequency function of X_1 , X_2 is $$p_{X_1 X_2}(x_1, x_2) = \sum_{x_3 \cdots x_m} p(x_1, x_2, \ldots, x_m)$$

What is the number of ways that n objects are grouped into r classes (types of outcomes) with n_i in the ith class, i=1, ..., r?

1st class: n_1 out of n; 2nd class: n_2 out of n-n_1;
… rth class: n_r out of $n-n_1-n_2-\ldots-n_{r-1}$

$$\frac{n!}{n_1!(n-n_1)!} \frac{(n-n_1)!}{(n-n_1-n_2)!n_2!} \cdots \frac{(n-n_1-n_2-\cdots-n_{r-1})!}{0!n_r!}$$

Multinomial coefficient: $$\binom{n}{n_1 n_2 \cdots n_r} = \frac{n!}{n_1!n_2!\cdots n_r!}$$

Example: Multinomial distribution

Each of n independent trials can result in one of r types of outcomes,
on each trial the probabilities of the r outcomes are $p_1, p_2, \ldots, p_r$.
N_i = total # of outcomes of type i in the n trials, $i=1, \ldots, r$.
(e.g. God is playing a dice…)

Any particular sequence of trials giving rise to $N_1=n_1, \ldots, N_r=n_r$ occurs with probability

$$p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

How many such sequences are there?

Example: Multinomial distribution

Each of n independent trials can result in one of r types of outcomes, on each trial the probabilities of the r outcomes are $p_1, p_2, ..., p_r$.
N_i = total # of outcomes of type i in the n trials, i=1, ..., r.
(e.g. God is playing a dice…)

Any particular sequence of trials giving rise to $N_1=n_1$, …, $N_r=n_r$ occurs with probability

$$p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

How many such sequences are there?
Equivalent question: What is the # of ways that n objects are grouped into r classes (types of outcomes) with n_i in the ith class, i=1, ..., r?

Joint frequency function: $$p(n_1, \ldots, n_r) = \binom{n}{n_1 \cdots n_r} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

Marginal distribution of N_i? -- Direct summation is daunting!

Example: Multinomial distribution

Each of n independent trials can result in one of r types of outcomes,
on each trial the probabilities of the r outcomes are $p_1, p_2, \ldots, p_r$.
N_i = total # of outcomes of type i in the n trials, i=1, ..., r.
(e.g. God is playing a dice…)

Any particular sequence of trials giving rise to $N_1=n_1, \ldots, N_r=n_r$ occurs with probability

$$p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

How many such sequences are there?
Equivalent question: What is the # of ways that n objects are grouped into r classes (types of outcomes) with n_i in the ith class, i=1, ..., r?

Joint frequency function: $$p(n_1, \ldots, n_r) = \binom{n}{n_1 \cdots n_r} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

Marginal distribution of N_i? -- Direct summation is daunting!
N_i can be interpreted as # of success in n trials, each of which has p_i of success.

Binomial random variable N_i renders $$p_{N_i}(n_i) = \binom{n}{n_i} p_i^{n_i} (1 - p_i)^{n-n_i}$$

Random variables: continuous

X, Y are continuous random variables with joint cdf *F* (*x, y*), joint pdf *f* (*x, y*).
For any "reasonable" 2-D set A

$$P((X, Y) \in A) = \iint_{\mathrm{A}} f(x, y)\, dy\, dx$$

In particular, if $A = \{(X, Y) | X \leq x \text{ and } Y \leq y\}$,

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dv\, du$$

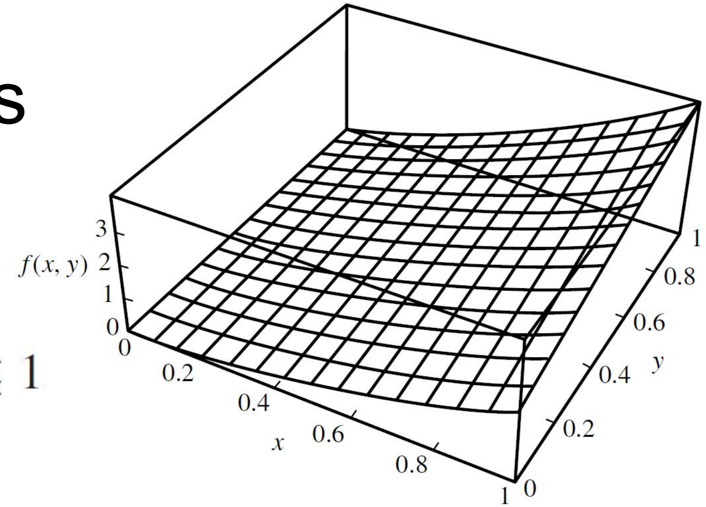
From the fundamental theorem of multivariable calculus, it follows that

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

Differential form is

$$P(x \leq X \leq x + dx, y \leq Y \leq y + dy) = f(x, y)\, dx\, dy$$

Random variables: continuous



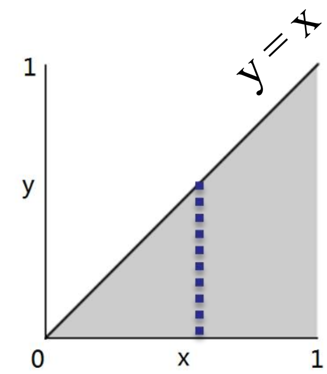
Example: Consider the bivariate density function

$$f(x, y) = \frac{12}{7}(x^2 + xy), \qquad 0 \le x \le 1, \qquad 0 \le y \le 1$$

$P\,(X>Y)$ can be found by integrating f over the set

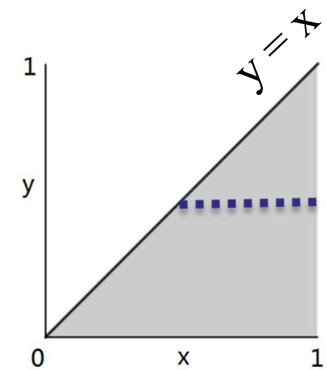
$$\{(x, y) | 0 \le y \le x \le 1\}$$

If we fix x first, y ranges from 0 to x, then x over [0, 1].



$$P(X > Y) = \frac{12}{7} \int_0^1 \boxed{\int_0^x (x^2 + xy)\, dy}\, dx = \frac{9}{14}$$

If we fix y first, x ranges from x to 1, then y over [0, 1].



$$P(X > Y) = \frac{12}{7} \int_0^1 \boxed{\int_y^1 (x^2 + xy)\, dx}\, dy$$

Random variables: continuous

The **marginal cdf** of X, or F_X, is

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \lim_{y \to \infty} F(x, y) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, y)\, dy\, du \end{aligned}$$

From this, it follows that the density function of X alone, known as the **marginal density** of X, is

$$f_X(x) = F'_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

Summing / integrating the joint frequency function over the other variable.

Random variables: continuous

The **marginal cdf** of X, or F_X, is

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \lim_{y \to \infty} F(x, y) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, y)\, dy\, du \end{aligned}$$

From this, it follows that the density function of X alone, known as the **marginal density** of X, is

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

Summing / integrating the joint frequency function over the other variable.

Example. Bivariate density function $f(x, y) = \frac{12}{7}(x^2 + xy),\ 0 \leq x, y \leq 1$

Marginal density of X: $f_X(x) = \frac{12}{7} \int_0^1 (x^2 + xy)\, dy = \frac{12}{7}\left(x^2 + \frac{x}{2}\right)$

Marginal density of Y: $f_Y(y) = \frac{12}{7}\left(\frac{1}{3} + y/2\right).$

For several jointly continuous random variables, there are marginal density functions of various dimensions. Say, for density function $f(x, y, z)$ of X, Y, Z,

1-d marginal distribution of X: $$f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y, z)\, dy\, dz$$

2-d marginal distribution of X and Y: $$f_{XY}(x, y) = \int_{-\infty}^{\infty} f(x, y, z)\, dz$$

Farlie-Morgenstern Family

If $F(x)$, $G(x)$ are 1-d cdfs, for any α with $|\alpha| \leq 1$, the following is a bivariate cumulative distribution function:

$$H(x, y) = F(x)G(y)\{1 + \alpha[1 - F(x)][1 - G(y)]\}$$

As $\lim_{x\to\infty} F(x) = \lim_{y\to\infty} F(y) = 1,$ the marginal distributions are

$$H(x, \infty) = F(x)$$
$$H(\infty, y) = G(y)$$

Marginals given, we may construct H as many as we want.

Farlie-Morgenstern Family

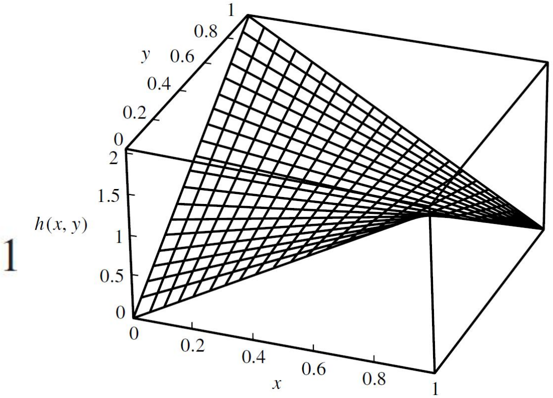
$$H(x, y) = F(x)G(y)\{1 + \alpha[1 - F(x)][1 - G(y)]\}$$

Example. Marginals are uniform on [0, 1], $F(x) = x$, $G(y) = y$, set $\alpha = -1$,

$$\begin{aligned} H(x, y) &= xy[1 - (1 - x)(1 - y)] \\ &= x^2y + y^2x - x^2y^2, \qquad 0 \le x \le 1, \qquad 0 \le y \le 1 \end{aligned}$$

Bivariate density

$$\begin{aligned} h(x, y) &= \frac{\partial^2}{\partial x \partial y} H(x, y) \\ &= 2x + 2y - 4xy, \qquad 0 \le x \le 1, \qquad 0 \le y \le 1 \end{aligned}$$

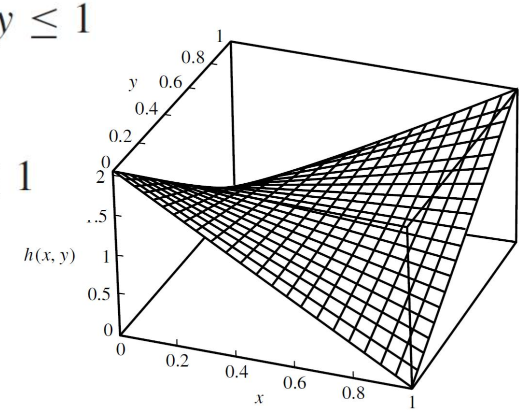


If $\alpha = 1$,

$$\begin{aligned} H(x, y) &= xy[1 + (1 - x)(1 - y)] \\ &= 2xy - x^2y - y^2x + x^2y^2, \qquad 0 \le x \le 1, \qquad 0 \le y \le 1 \end{aligned}$$

The density is

$$h(x, y) = 2 - 2x - 2y + 4xy, \qquad 0 \le x \le 1, \qquad 0 \le y \le 1$$



Farlie-Morgenstern Family are copulas

– a **copula（连接函数）** is a joint cdf that has **uniform marginal distributions**.

Note $P(U \le u) = C(u, 1) = u$ and $C(1, v) = v$

(summing up all v values over pdf is equivalent to let v approach infinity in cdf)

The density is

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v) \ge 0$$

Suppose X, Y have cdfs $F_X(x)$, $F_Y(y)$, then $U=F_X(x)$, $V=F_Y(y)$ are uniform random variables. For a copula $C(u, v)$, consider the joint cdf

$$F_{XY}(x, y) = C(F_X(x), F_Y(y))$$

Since $C(F_X(x), 1)=F_X(x)$, marginal cdfs of F_{XY} are $F_X(x)$, $F_Y(y)$. The pdf is

$$f_{XY}(x, y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y)$$

From ingredients of two marginal distributions and any copula, a joint distribution can be constructed.

Dependence between the random variables are captured in the copula.

PROPOSITION **C**

Let $Z = F(X)$; then Z has a uniform distribution on [0, 1].

Proof

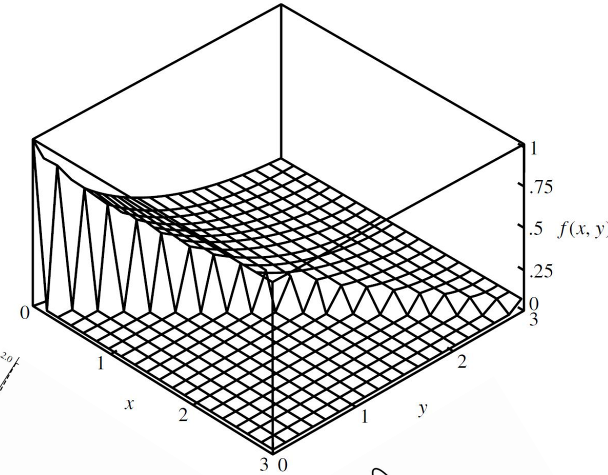
$$P(Z \le z) = P(F(X) \le z) = P(X \le F^{-1}(z)) = F(F^{-1}(z)) = z$$

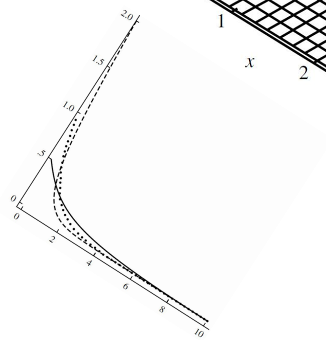
This is the uniform cdf. ■

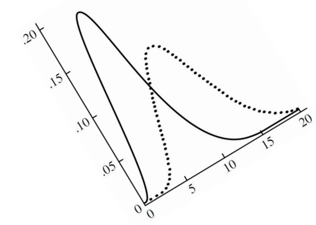
Example 1: The density is

$$f(x, y) = \begin{cases} \lambda^2 e^{-\lambda y}, & 0 \le x \le y, \lambda > 0 \\ 0, & \text{elsewhere} \end{cases}$$

Find marginal distributions.







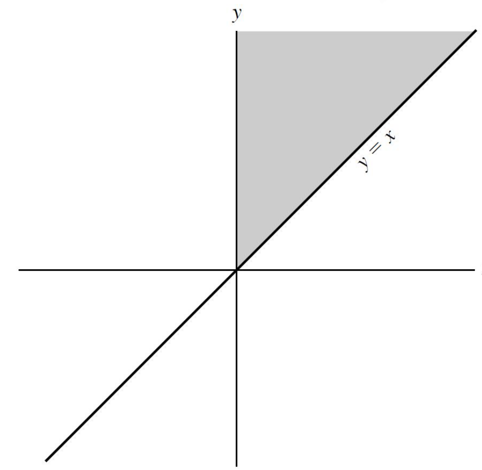
$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$
$$= \int_x^{\infty} \lambda^2 e^{-\lambda y} dy = \lambda e^{-\lambda x}, \qquad x \ge 0$$

- exponential.

$$f_Y(y) = \int_0^y \lambda^2 e^{-\lambda y} dx = \lambda^2 y e^{-\lambda y}, \qquad y \ge 0$$

- Gamma distribution. Recall

$$g(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}, \qquad t \ge 0$$



Picture showing where pdf is nonzero, to aid in determining the limits of integration

Example 2: A point is chosen randomly in a disk of radius 1.

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & \text{if } x^2 + y^2 \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

R is distance from origin, its cdf, pdf are

$$F_R(r) = P(R \leq r) = \frac{\pi r^2}{\pi} = r^2 \qquad f_R(r) = 2r, 0 \leq r \leq 1.$$

Marginal density of *x*?

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y)\, dy \\ &= \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy \\ &= \frac{2}{\pi}\sqrt{1 - x^2}, \qquad -1 \leq x \leq 1 \end{aligned}$$

What about *y*? Symmetry!

$$f_Y(y) = \frac{2}{\pi}\sqrt{1 - y^2}, \qquad -1 \leq y \leq 1$$

Example 3: Bivariate normal density

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right)$$

Five parameters

$$-\infty < \mu_X < \infty \qquad -\infty < \mu_Y < \infty$$

$$\sigma_X > 0 \qquad \sigma_Y > 0$$

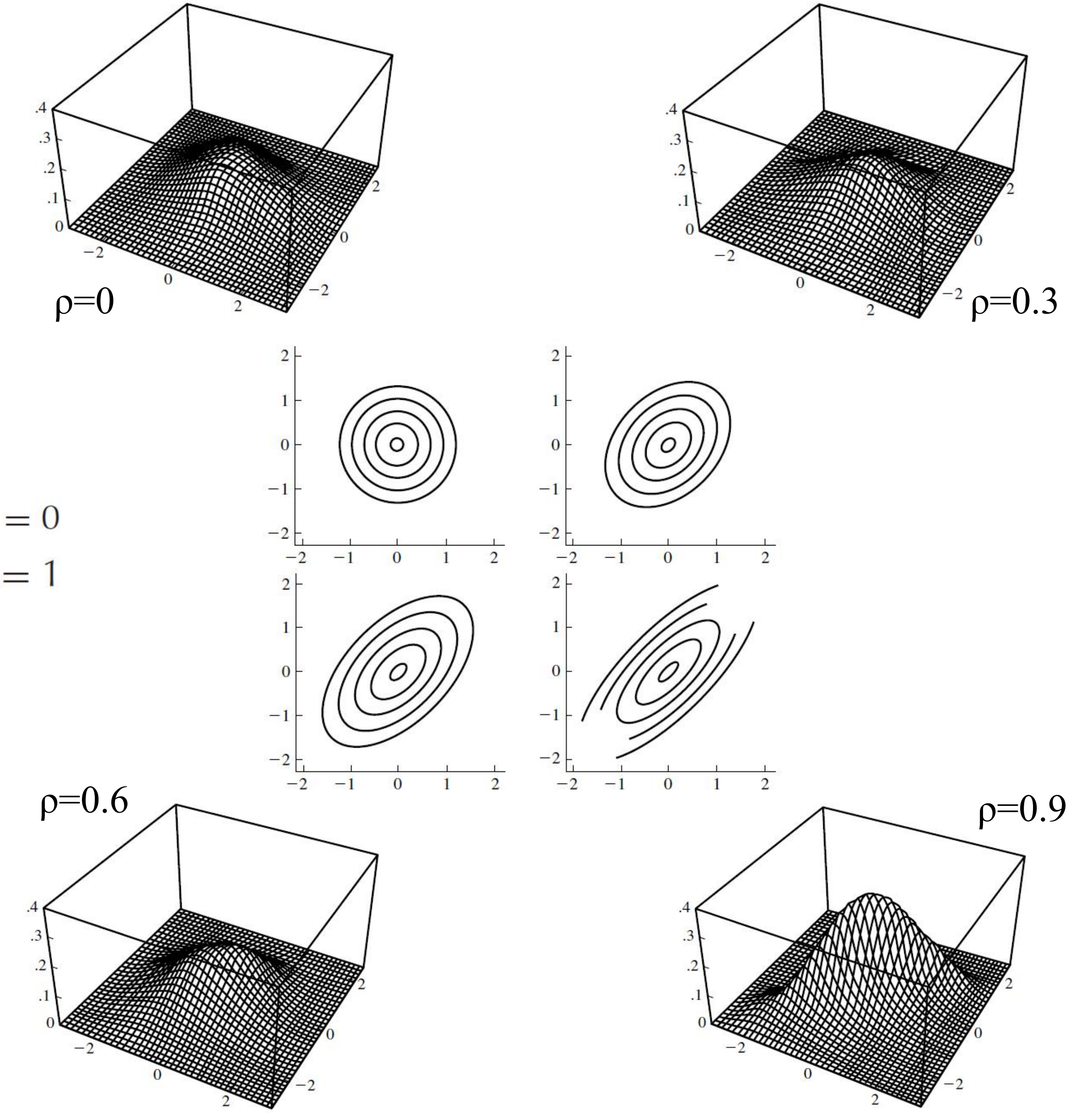
$$-1 < \rho < 1$$

What does its contour lines look like?

$$\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} = \text{constant}$$

-- An ellipse centered at (μ_X, μ_Y). $\rho \neq 0$, tilted; $\rho = 0$, untilted.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$$



$\mu_X = \mu_Y = 0$

$\sigma_X = \sigma_Y = 1$

Example 3: Bivariate normal density

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right)$$

Marginal density of X is

$$u = (x-\mu_X)/\sigma_X$$

$$v = (y-\mu_Y)/\sigma_Y$$

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

$$= \frac{1}{2\pi\sigma_X\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2(1-\rho^2)}(u^2 + v^2 - 2\rho uv)\right] dv$$

Completing the square, $(v-\rho u)^2 + u^2(1-\rho^2)$

$$= \frac{1}{2\pi\sigma_X\sqrt{1-\rho^2}} e^{-u^2/2} \underbrace{\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2(1-\rho^2)}(v-\rho u)^2\right] dv}_{\text{Gaussian}}$$

$$= \frac{1}{\sigma_X\sqrt{2\pi}} e^{-(1/2)\left[(x-\mu_X)^2/\sigma_X^2\right]}$$

If $X, Y \sim$ bivariate Gaussian distribution, their marginal distributions are Gaussian.

Inverse question?

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$$

Independently random variables

DEFINITION

Random variables $X_1, X_2, \ldots, X_n$ are said to be *independent* if their joint cdf factors into the product of their marginal cdf's:

$$F(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n)$$

for all $x_1, x_2, \ldots, x_n$. ■

If g and h are function, then Z=g(X) and W=h(Y) are also independent.

Example: Farlie-Morgenstern family, when $\alpha=0$, X, Y are independent.

$$H(x, y) = F(x)G(y)\{1 + \alpha[1 - F(x)][1 - G(y)]\}$$

Example: Bivariate normal distribution, when $\rho=0$, X, Y are independent.

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right)$$

Independently random variables

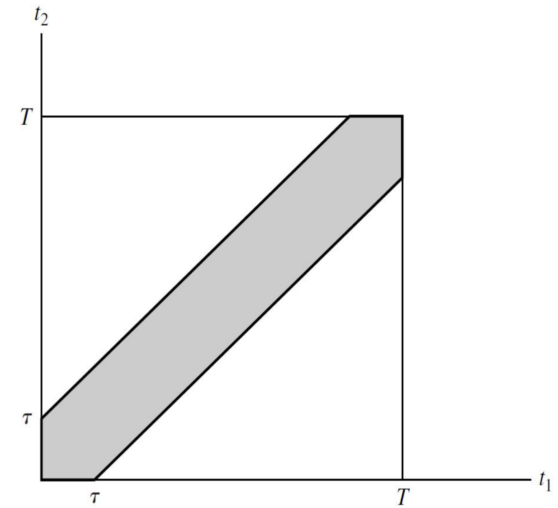
Example.

Suppose that a node in a communications network has the property that if two packets of information arrive within time τ of each other, they "collide" and then have to be retransmitted. If the times of arrival of the two packets are independent and uniform on $[0, T]$, what is the probability that they collide?

The times of arrival of two packets, T_1, T_2 are independent, both uniformly distributed on $[0, T]$, then joint pdf is

$$f(t_1, t_2) = \frac{1}{T^2}$$

We need $|t_1 - t_2| < \tau$, or $t_1 - \tau$ (or 0) $< t_2 < t_1 + \tau$



P(collision)

= pdf integrated over shaded area $T^2 - (T - \tau)^2$

$= 1 - (1 - \tau/T)^2.$

Conditional distributions

Conditional distributions: discrete

X, Y are discrete, the conditional probability that $X=x_i$ given that $Y=y_i$ is

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$
$$= \frac{p_{XY}(x_i, y_j)}{p_Y(y_j)}$$

We denote it as $p_{X|Y}(x|y)$. If *X, Y* are independent, $= p_Y(y)$

Reexpressing it,

$$p_{XY}(x, y) = p_{X|Y}(x|y) p_Y(y)$$

Summing both sides over all *y* values, law of total probability,

$$p_X(x) = \sum_y p_{X|Y}(x|y) p_Y(y)$$

Conditional distributions: discrete

Example.

Suppose that a particle counter is imperfect and independently detects each incoming particle with probability p. If the distribution of the number of incoming particles in a unit of time is a Poisson distribution with parameter λ, what is the distribution of the number of counted particles?

Make a guess!!

Conditional distributions: discrete

Example.

Suppose that a particle counter is imperfect and independently detects each incoming particle with probability p. If the distribution of the number of incoming particles in a unit of time is a Poisson distribution with parameter λ, what is the distribution of the number of counted particles?

$$
\begin{aligned}
P(X=k) &= \sum_{n=0}^{\infty} P(N=n)P(X=k|N=n) \\
&= \sum_{n=k}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{n=k}^{\infty} \lambda^{n-k} \frac{(1-p)^{n-k}}{(n-k)!} \\
&= \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j (1-p)^j}{j!} \\
&= \frac{(\lambda p)^k}{k!} e^{-\lambda} e^{\lambda(1-p)} \\
&= \frac{(\lambda p)^k}{k!} e^{-\lambda p}
\end{aligned}
$$

N=true # of particles
X=# of detected particles

Given $N=n$, X is binomial

Denote $j=n-k$

Still Poisson, but parameter is now λp.

Conditional distributions: continuous

The conditional probability of Y given X is defined to be

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

$$\Updownarrow$$

$$P(y \le Y \le y + dy | x \le X \le x + dx) = \frac{f_{XY}(x, y)\, dx\, dy}{f_X(x)\, dx} = \frac{f_{XY}(x, y)}{f_X(x)} dy$$

Reexpressing it, $f_{XY}(x, y) = f_{Y|X}(y|x) f_X(x)$

Integrating both sides over x, the marginal density of Y is

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

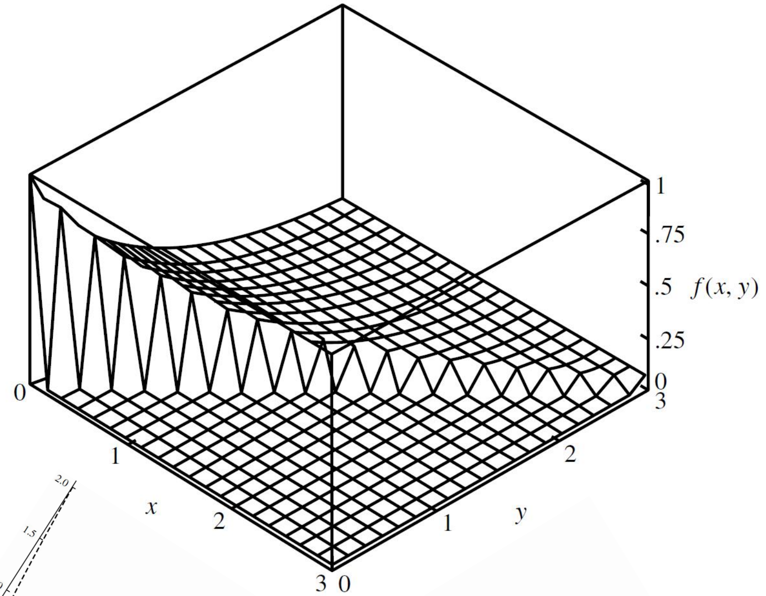
Law of total probability for the continuous case.

Previous example

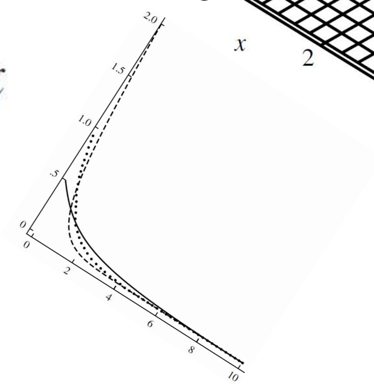
$$f_{XY}(x, y) = \lambda^2 e^{-\lambda y}, \qquad 0 \le x \le y$$

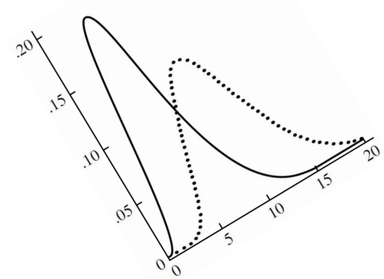
$$f_X(x) = \lambda e^{-\lambda x}, \qquad x \ge 0$$

$$f_Y(y) = \lambda^2 y e^{-\lambda y}, \qquad y \ge 0$$



$$f_{Y|X}(y|x) = \frac{\lambda^2 e^{-\lambda y}}{\lambda e^{-\lambda x}} = \lambda e^{-\lambda(y-x)}, \qquad y \ge x$$



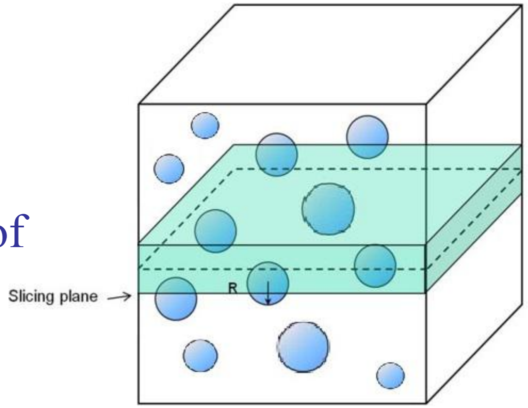


$$f_{X|Y}(x|y) = \frac{\lambda^2 e^{-\lambda y}}{\lambda^2 y e^{-\lambda y}} = \frac{1}{y}, \qquad 0 \le x \le y$$

Three interesting examples -- 1. Stereology (体视学)

In metallography (金相学）and other applications of quantitative microscopy, aspects of a 3-d structure are deduced from studying 2-d cross sections, where statistics play an important role (DeHoff & Rhines 1968).

Spherical particles are dispersed in a medium (e.g. grains in a metal); the density function of the radii of the spheres is denoted as $f_R(r)$. When the medium is sliced, 2-d, circular cross sections of the spheres are observed. We denote the pdf of the radii of these circles as $f_X(x)$. Their relationship?

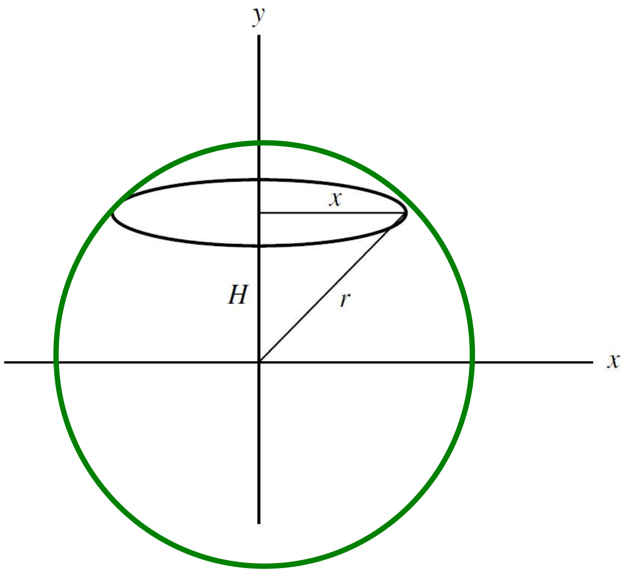


From the viewpoint of a single sphere, the slicing plane is chosen at random. Fix $R=r$, find conditional density $f_{X|R}(x|r)$.

H is uniformly distributed over $[0, r]$, $X = \sqrt{r^2 - H^2}$.

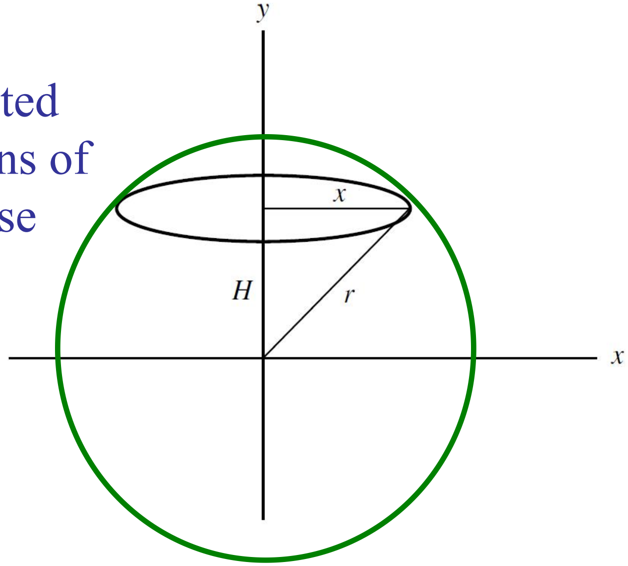
$$
\begin{aligned}
F_{X|R}(x|r) &= P(X \le x) \\
&= P(\sqrt{r^2 - H^2} \le x) \\
&= P(H \ge \sqrt{r^2 - x^2}) \\
&= 1 - \frac{\sqrt{r^2 - x^2}}{r}, \qquad 0 \le x \le r
\end{aligned}
$$

r fixed!



Three interesting examples -- 1. Stereology (体视学)

Spherical particles are dispersed in a medium (e.g. grains in a metal); the density function of the radii of the spheres is denoted as $f_R(r)$. When the medium is sliced, 2-d, circular cross sections of the spheres are observed. We denote the pdf of the radii of these circles as $f_X(x)$. Their relationship?



Differentiating it w.r.t. x,

$$f_{X|R}(x|r) = \frac{x}{r\sqrt{r^2 - x^2}}, \qquad 0 \le x \le r$$

The law of total probability gives marginal density of X,

$$\underset{observable}{f_X(x)} = \int_{-\infty}^{\infty} f_{X|R}(x|r) f_R(r)\, dr$$

$$= \int_{\underline{x}}^{\infty} \frac{x}{r\sqrt{r^2 - x^2}} \overset{wanted}{f_R(r)}\, dr$$

Abel's equation / transform

- bridging 2-D projection and 3-D spherical geometry.
- Useful in astronomy!! (e.g. galactic astronomy, cosmology…)

Niels Henrik Abel (1802-1829),
Norwegian mathematician,
Professor-to-be at Univ. of Berlin

Three interesting examples -- 2. Rejection method

Lec #2: To generate random variables with cdf F, apply F^{-1} to uniform random variables.

But what if F^{-1} does not have a closed form?

Suppose pdf f is defined on $[a, b]$, let M be a function $M(x) \geq f(x)$ and let m be a pdf,

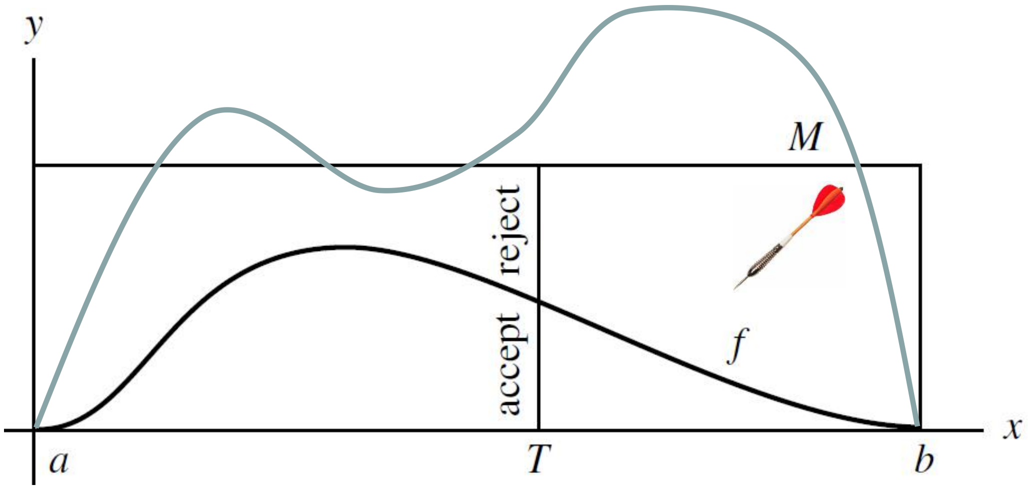
$$m(x) = \frac{M(x)}{\int_a^b M(x)\,dx}$$

***Step* 1:** Generate T with pdf m.

***Step* 2:** Generate U, uniform on [0, 1] and independent of T.

- if $M(T)\cdot U \leq f(T)$, then let $X=T$ (accept T);
- otherwise, reject T (go to step 1).

Idea is to choose M so that it is easy to generate random variables from m.



Three interesting examples -- 2. Rejection method

Justification: What is *X*? *T* but only if it is accepted.

Why not the opposite?

$$P(x \le X \le x+dx) = P(x \le T \le x+dx \mid \text{accept})$$
$$= \frac{P(x \le T \le x+dx \ \text{ and accept})}{P(\text{accept})}$$
$$= \frac{P(\text{accept}|x \le T \le x+dx)P(x \le T \le x+dx)}{P(\text{accept})}$$

$$m(x)\ dx$$

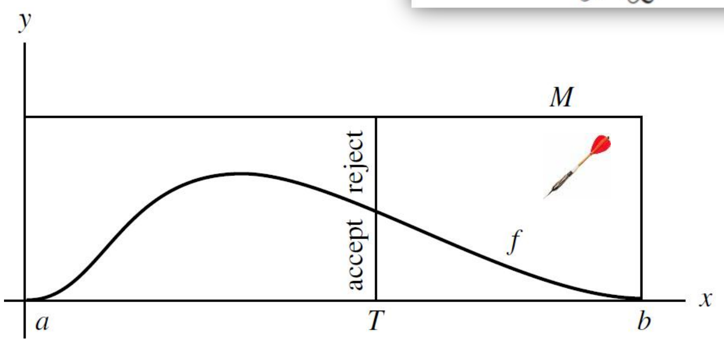
$$P(\text{accept}|x \le T \le x+dx) = P(U \le f(x)/M(x)) = \frac{f(x)}{M(x)}$$

$$P(\text{accept}) = P(U \le f(T)/M(T))$$
$$= \int_a^b \frac{f(t)}{M(t)} m(t)\ dt = \frac{1}{\int_a^b M(t)\ dt}$$

Law of total probability

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

For any given $T=t$, probability of $U \le f(T)/M(T)$ is $f(t)/M(t)$. Then sum over all t values.



Collecting everything, LHS = $f(x)\ dx$.

Three interesting examples -- 3. Bayesian Inference

A freshly minted coin has a certain probability of coming up heads if it is spun on its edge (may not be ½). Say, you spin it n times and see X heads. What has been learned about the chance Θ it comes up heads?

Totally ignorant about it, we might represent our knowledge by a uniform density on [0, 1], the **prior density**

$$f_\Theta(\theta) = 1, \quad 0 \le \theta \le 1.$$

Given a value θ, X follows binomial distribution

$$f_{X|\Theta}(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \qquad x = 0, 1, \ldots, n$$

The joint discrete/continuous pdf

$$f_{\Theta,X}(\theta, x) = f_{X|\Theta}(x|\theta) f_\Theta(\theta)$$

$$= \binom{n}{x}\theta^x(1-\theta)^{n-x}, \qquad x = 0, 1, \ldots, n, \quad 0 \le \theta \le 1$$

Marginal density of X by integrating over θ,

$$f_X(x) = \int_0^1 \binom{n}{x}\theta^x(1-\theta)^{n-x}\,d\theta$$

$$f_{\Theta|X}(\theta|x) = \frac{f_{\Theta,X}(\theta, x)}{f_X(x)}$$

Three interesting examples -- 3. Bayesian Inference

Using Gamma and Beta functions (I skip the tedious maths here),

$$f_X(x) = \int_0^1 \binom{n}{x} \theta^x (1-\theta)^{n-x} d\theta = \ldots = \frac{1}{n+1}, \qquad x = 0, 1, \ldots, n$$

(If our prior on θ is uniform, each outcome of X is a priori equally likely).

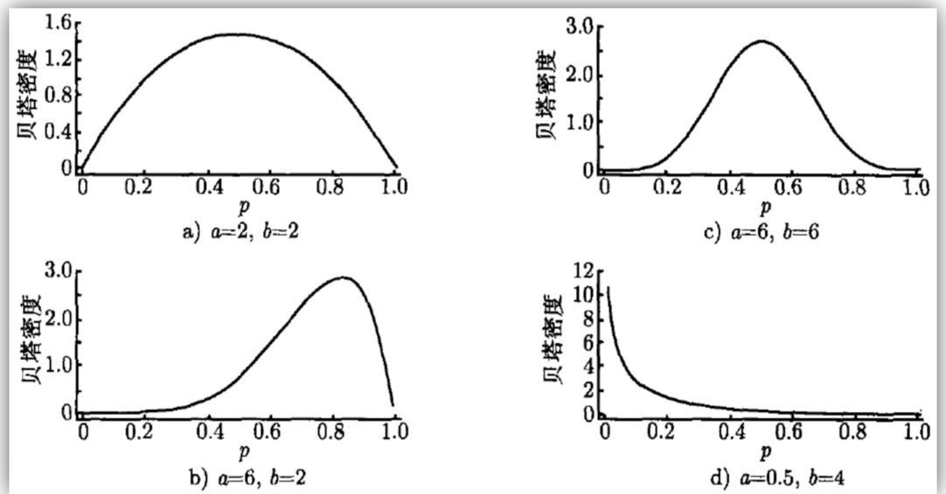
<u>Having observed X=x, what's our knowledge on Θ?</u>
Quantified by the posterior density of θ given x,

$$f_{\Theta|X}(\theta|x) = \frac{f_{\Theta,X}(\theta, x)}{f_X(x)} = (n+1)\binom{n}{x}\theta^x(1-\theta)^{n-x}$$

$$= \ldots = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x}$$

Posterior density is **Beta density** *with* $a=x+1$, $b=n-x+1$.

$$g(u) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1}(1-u)^{b-1}, \qquad 0 \le u \le 1$$



Three interesting examples -- 3. Bayesian Inference

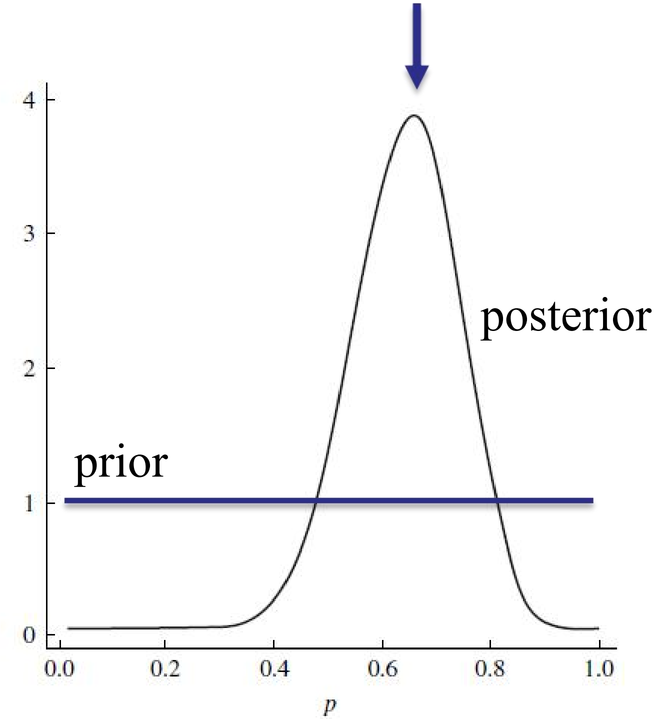
$$f_{\Theta|X}(\theta|x) = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x}$$

Posterior is ***Beta density****, a=x+1, b=n−x+1.*

Suppose we spin a coin n=20 times, finding x=13 heads.
a=x+1=14, b=n−x+1=8, see figure.

- Extremely unlikely that θ<0.25
- Likely θ>0.5 (91% probability)

Question: where does it peak?



Three interesting examples -- 3. Bayesian Inference

$$f_{\Theta|X}(\theta|x) = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x}$$

*Posterior is **Beta density**, $a=x+1$, $b=n-x+1$.*

Suppose we spin a coin $n=20$ times, finding $x=13$ heads.
$a=x+1=14$, $b=n-x+1=8$, see figure.

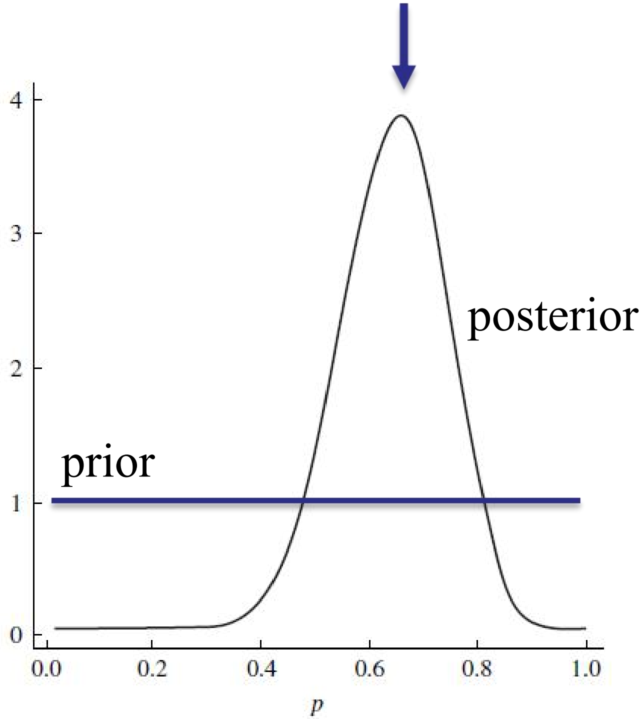
- Extremely unlikely that $\theta<0.25$
- Likely $\theta>0.5$ (91% probability)

Question: where does it peak?

$\theta^{13}(1-\theta)^7$ peaks at 13/20. Intuitively correct.

Analog questions in astronomy are common.

*We observe a fraction, deduce the real fraction (**uncertainty directly given by Bayesian inference**).*



Functions of jointly distributed random variables

PROPOSITION **B**

Let X be a continuous random variable with density $f(x)$ and let $Y = g(X)$ where g is a differentiable, strictly monotonic function on some interval I. Suppose that $f(x) = 0$ if x is not in I. Then Y has the density function

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

for y such that $y = g(x)$ for some x, and $f_Y(y) = 0$ if $y \neq g(x)$ for any x in I. Here g^{-1} is the inverse function of g; that is, $g^{-1}(y) = x$ if $y = g(x)$. ■

Special case A: sums

X, Y are discrete random variables, taking integer values and having joint frequency function $p(x, y)$. What is the frequency function of $Z=X+Y$?

Whenever $X=x$ and $Y=z-x$, $Z=z$, so the probability that $Z=z$ is the sum over all x of these joint probabilities,

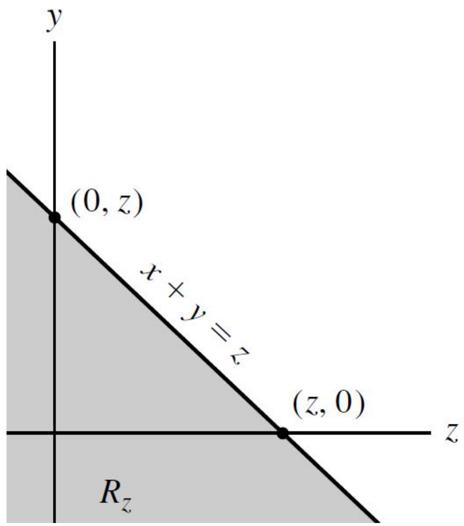
$$p_Z(z) = \sum_{x=-\infty}^{\infty} p(x, z-x)$$

If X and Y are independent,

$$p_Z(z) = \sum_{x=-\infty}^{\infty} p_X(x) p_Y(z-x)$$

- **convolution** of the sequences p_X and p_Y.

Special case A: sums



Continuous case

Find cdf of Z first!

$$F_Z(z) = \iint_{R_z} f(x, y)\, dx\, dy$$

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

Differentiating it, using the rule of chains,

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x)\, dx$$

If X and Y are independent,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$$

- convolution of the functions f_X and f_Y.

Convolution is everywhere in astronomy:

- Smooth an image with a PSF
- Deriving the star formation history
- combination of multiple effects
- …

Special case A: sums

Continuous case

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$$

Example: The lifetime of a component is exponentially distributed. We have an identical and independent backup component.

Lifetime of the system $S = T_1 + T_2$, its pdf is

$$f_S(s) = \int_0^s \lambda e^{-\lambda t} \lambda e^{-\lambda(s-t)} dt$$

Beyond the limits of integration, both components have 0 density.

$$f_S(s) = \lambda^2 \int_0^s e^{-\lambda s}\, dt$$

$$= \lambda^2 s e^{-\lambda s}$$

$$g(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}, \qquad t \geq 0$$

Gamma distribution with parameters 2, λ.

Special case B: quotients

Find cdf of $Z=Y/X$ first, again.

$F_Z(z) = P(Z \leq z)$ is the probability of the set of (x, y) such that $y/x \leq z$.
If $x > 0$, it is the set $y \leq xz$; if $x < 0$, it is the set $y \geq xz$.

$$F_Z(z) = \int_{-\infty}^{0} \int_{xz}^{\infty} f(x, y)\, dy\, dx + \int_{0}^{\infty} \int_{-\infty}^{xz} f(x, y)\, dy\, dx$$

Differentiating it, using the rule of chains, insert $y=xv$ (instead of xz),

$$F_Z(z) = \int_{-\infty}^{0} \int_{z}^{-\infty} xf(x, xv)\, dv\, dx + \int_{0}^{\infty} \int_{-\infty}^{z} xf(x, xv)\, dv\, dx$$

$$= \int_{-\infty}^{0} \int_{-\infty}^{z} (-x) f(x, xv)\, dv\, dx + \int_{0}^{\infty} \int_{-\infty}^{z} xf(x, xv)\, dv\, dx$$

$$= \int_{-\infty}^{z} \int_{-\infty}^{\infty} |x| f(x, xv)\, dx\, dv$$

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f(x, xz)\, dx$$

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz)\, dx$$

If X, Y are independent.

Special case B: quotients

Example: How is the ratio of two independent Gaussians distributed?

Consider standard normal distribution, $Z=Y/X$, then

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{|x|}{2\pi} e^{-x^2/2} e^{-x^2 z^2/2}\, dx$$

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f(x, xz)\, dx$$

Integrand is even,

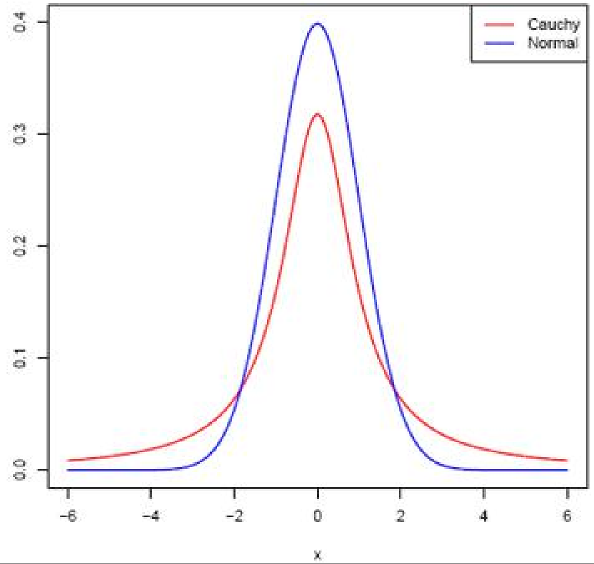
$$f_Z(z) = \frac{1}{\pi} \int_0^{\infty} x e^{-x^2((z^2+1)/2)}\, dx$$

$$u = x^2$$

$$= \frac{1}{2\pi} \int_0^{\infty} e^{-u((z^2+1)/2)}\, du$$

$$\lambda = (z^2+1)/2$$

$$\int_0^{\infty} \lambda \exp(-\lambda x)\, dx = 1$$



$$f_Z(z) = \frac{1}{\pi(z^2+1)}, \qquad -\infty < z < \infty$$

Cauchy density decreases slower than Gaussians.

The general case

Example: X, Y are independent standard normal random variables, joint pdf is

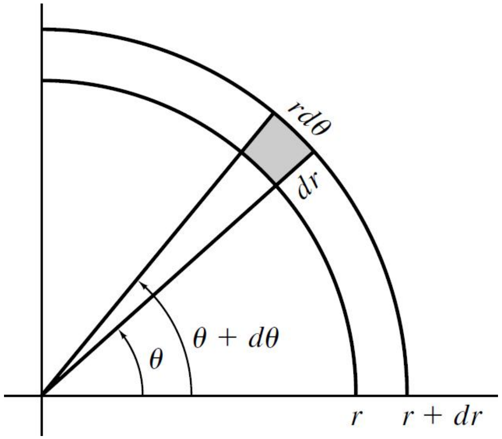
$$f_{XY}(x, y) = \frac{1}{2\pi} e^{-(x^2/2)-(y^2/2)}$$

Changing to polar coordinates,

$$R = \sqrt{X^2 + Y^2}$$

$$\Theta = \begin{cases} \tan^{-1}\left(\frac{Y}{X}\right), & \text{if } X > 0 \\ \tan^{-1}\left(\frac{Y}{X}\right) + \pi, & \text{if } X < 0 \\ \frac{\pi}{2}\operatorname{sgn}(Y), & \text{if } X = 0, Y \neq 0 \\ 0, & \text{if } X = 0, Y = 0 \end{cases}$$

$$X = R\cos\Theta$$
$$Y = R\sin\Theta$$



The general case

R, Θ have joint distribution

$$f_{R\Theta}(r,\theta)\,dr\,d\theta = P(r \le R \le r+dr, \theta \le \Theta \le \theta + d\theta)$$
$$= f_{XY}(r\cos\theta, r\sin\theta) r\,dr\,d\theta$$

$$f_{R\Theta}(r,\theta) = r f_{XY}(r\cos\theta, r\sin\theta)$$

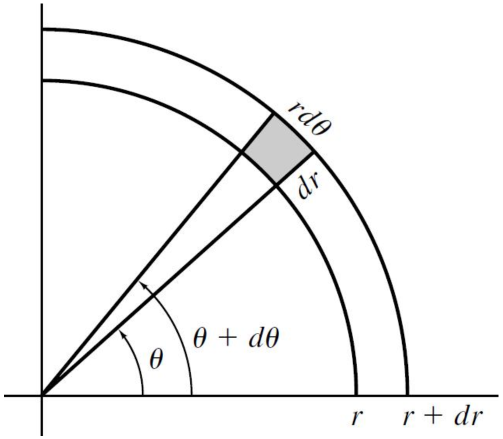
$$f_{R\Theta}(r,\theta) = \frac{r}{2\pi} e^{[-(r^2\cos^2\theta)/2-(r^2\sin^2\theta)/2]}$$
$$= \frac{1}{2\pi} r e^{-r^2/2}$$

$$f_{XY}(x,y) = \frac{1}{2\pi} e^{-(x^2/2)-(y^2/2)}$$

Joint density implies that R and Θ are independent variables, Θ is uniform on $[0, 2\pi]$, R has the density

$$f_R(r) = re^{-r^2/2}, \qquad r \ge 0$$

Rayleigh density!



The general case

Under the assumptions just stated, the joint density of U and V is

$$f_{UV}(u, v) = f_{XY}(h_1(u, v), h_2(u, v))|J^{-1}(h_1(u, v), h_2(u, v))|$$

for (u, v) such that $u = g_1(x, y)$ and $v = g_2(x, y)$ for some (x, y) and 0 elsewhere. ■

If $X_1, \ldots, X_n$ have the joint density function $f_{X_1 \cdots X_n}$ and

$$Y_i = g_i(X_1, \ldots, X_n), \qquad i = 1, \ldots, n$$

$$X_i = h_i(Y_1, \ldots, Y_n), \qquad i = 1, \ldots, n$$

and if $J(x_1, \ldots, x_n)$ is the determinant of the matrix with the ij entry $\partial g_i / \partial x_j$, then the joint density of $Y_1, \ldots, Y_n$ is

$$f_{Y_1 \cdots Y_n}(y_1, \ldots, y_n) = f_{X_1 \cdots X_n}(x_1, \ldots, x_n)|J^{-1}(x_1, \ldots, x_n)|$$

wherein each x_i is expressed in terms of the y's; $x_i = h_i(y_1, \ldots, y_n)$.