Random Variables

$X \sim F_X(x)$: a random variable X distributed with CDF F_X.

Any function $Y = g(X)$ is also a random variable.

If both X, and Y are continuous random variables, can we find a simple way to characterize F_Y and f_Y (the CDF and PDF of Y), based on the CDF and PDF of X?

$\blacksquare\blacksquare\blacksquare$

For the CDF:

$$
\begin{aligned}
F_Y(y) &= P_Y(Y \leq y) \\
&= P_Y(g(X) \leq y) \\
&= P_X(x \in \mathcal{X} : g(X) \leq y) \text{ } (\mathcal{X} \text{ is sample space for } X) \\
&= \int_{\{x \in \mathcal{X} : g(X) \leq y\}} f_X(s)ds.
\end{aligned}
$$

PDF: $f_Y(y) = F'_y(y)$

Caution: need to consider support of y.

Consider several examples:

1. $X \sim U[-1, 1]$ and $y = \exp(x)$

   That is:

$$
\begin{aligned}
f_X(x) &= \begin{cases} \frac{1}{2} & \text{if } x \in [-1, 1] \\ 0 & \text{otherwise} \end{cases} \\
F_X(x) &= \frac{1}{2} + \frac{1}{2}x, \text{ for } x \in [-1, 1].
\end{aligned}
$$

$$
\begin{aligned}
F_Y(y) &= Prob(\exp(X) \leq y) \\
&= Prob(X \leq \log y) \\
&= F_X(\log y) = \frac{1}{2} + \frac{1}{2}\log y, \quad \text{for } y \in [\frac{1}{e}, e].
\end{aligned}
$$

   Be careful about the bounds of the support!

$$
\begin{aligned}
f_Y(y) &= \frac{\partial}{\partial y} F_Y(y) \\
&= f_X(\log y)\frac{1}{y} = \frac{1}{2y}, \text{ for } y \in [\frac{1}{e}, e].
\end{aligned}
$$

2. $X \sim U[-1,1]$ and $Y = X^2$

$$\begin{aligned}
F_Y(y) &= Prob(X^2 \le y) \\
&= Prob(-\sqrt{y} \le X \le \sqrt{y}) \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\
&= 2F_X(\sqrt{y}) - 1, \text{ by symmetry: } F_X(-\sqrt{y}) = 1 - F_X(\sqrt{y}).
\end{aligned}$$

$$\begin{aligned}
f_Y(y) &= \frac{\partial}{\partial y} F_Y(y) \\
&= 2f_X(\sqrt{y})\frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}}, \text{ for } y \in [0,1].
\end{aligned}$$

■■■

As the first example above showed, it's easy to derive the CDF and PDF of Y when $g(\cdot)$ is a strictly monotonic function:

Theorems 2.1.3, 2.1.5: When $g(\cdot)$ is a strictly increasing function, then

$$\begin{aligned}
F_Y(y) &= \int_{-\infty}^{g^{-1}(y)} f_X(x)dx = F_X(g^{-1}(y)) \\
f_Y(y) &= f_X(g^{-1}(y))\frac{\partial}{\partial y} g^{-1}(y) \text{ using chain rule.}
\end{aligned}$$

Note: by the inverse function theorem,

$$\frac{\partial}{\partial y} g^{-1}(y) = 1/\left[g'(x)\right]|_{x=g^{-1}(y)}.$$

When $g(\cdot)$ is a strictly decreasing function, then

$$\begin{aligned}
F_Y(y) &= \int_{g^{-1}(y)}^{\infty} f_X(x)dx = 1 - F_X(g^{-1}(y)) \\
f_Y(y) &= -f_X(g^{-1}(y))\frac{\partial}{\partial y} g^{-1}(y) \text{ using chain rule.}
\end{aligned}$$

These are the *change of variables* formulas for transformations of univariate random variables.

transformations.

■■■

Here is a special case of a transformation:

Thm 2.1.10: Let X have a continuous CDF $F_X(\cdot)$ and define the random variable $Y = F_X(X)$. Then $Y \sim U[0,1]$, i.e., $F_Y(y) = y$, for $y \in [0,1]$.

■■■

Expected value (Definition 2.2.1): The expected value, or mean, of a random variable $g(X)$ is

$$Eg(X) = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } X \text{ continuous} \\ \sum_{x \in \mathcal{X}} g(x) P(X = x) & \text{if } X \text{ discrete} \end{cases}$$

provided that the integral or the sum exists

The expectation is a *linear operator* (just like integration): so that

$$E\left[\alpha * \sum_{i=1}^{n} g_i(X) + b\right] = \alpha * \sum_{i=1}^{n} Eg_i(X) + b.$$

Note: Expectation is a *population average*, i.e., you average values of the random variable $g(X)$ weighting by the population density $f_X(x)$.

A statistical experiment yields a *finite sample* of observations $X_1, X_2, \ldots, X_n \sim F_X$. From a *finite sample*, you can never know the expectation. From these sample observations, we can calculate sample avg. $\bar{X}_n \equiv \frac{1}{n}\sum_i X_i$. In general: $\bar{X}_n \neq EX$. But under some conditions, as $n \to \infty$, then $\bar{X}_n \to EX$ in some sense (which we discuss later).

■■■

Expected value is commonly used measure of "central tendency" of a random variable X.

Example: But mean may not exist: Cauchy random variable with density $f(x) = \frac{1}{\pi(1+x^2)}$ for $x \in (-\infty, \infty)$. Note that

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)} dx &= \int_{-\infty}^{0} \frac{x}{\pi(1+x^2)} dx + \int_{0}^{\infty} \frac{x}{\pi(1+x^2)} dx \\ &= \lim_{a \to -\infty}^{0} \int_{a}^{0} \frac{x}{\pi(1+x^2)} dx + \lim_{b \to \infty} \int_{0}^{b} \frac{x}{\pi(1+x^2)} dx \\ &= \lim_{a \to -\infty}^{0} \frac{1}{2\pi} [\log(1+x^2)]_a^0 + \lim_{b \to \infty} \frac{1}{2\pi} [\log(1+x^2)]_0^b \\ &= -\infty + \infty \quad \text{undefined} \end{aligned}$$

■

Other measures:

1. Median: $\text{med}(X) = m$ such that $F_X(x) = 0.5$. Robust to outliers, and has nice invariance property: for $Y = g(X)$ and $g(\cdot)$ monotonic increasing, then $\text{med}(Y) = g(\text{med}(X))$.
2. Mode: $\text{Mode}(X) = \max_x f_X(x)$.

■■■

Moments: important class of expectations

For each integer n, the n-th (uncentred) moment of $X \sim F_X(\cdot)$ is $\mu'_n \equiv EX^n$.

The n-th centred moment is $\mu_n \equiv E(X - \mu)^n = E(X - EX)^n$. (It is centred around the mean EX.)

For $n = 2$: $\mu_2 = E(X - EX)^2$ is the *Variance* of X. $\sqrt{\mu_2}$ is the *standard deviation.*

Important formulas:

- $Var(aX + b) = a^2 VarX$ (variance is not a linear operation)
- $VarX = E(X^2) - (EX)^2$: alternative formula for the variance

■■■

Characteristic function:

The characteristic function of a random variable x, defined as

$$\phi_x(t) = E_x \exp(itx) = \int_{-\infty}^{+\infty} \exp(itx) f(x) dx$$

where $f(x)$ is the density for x.

This is also called the "Fourier transform".

Features of characteristic function:

- The CF always exists. This follows from the equality $e^{itx} = \cos(tx) + i \cdot \sin(tx)$. Note that the modulus $|e^{itx}| = \sqrt{\cos^2(x) + \sin^2(x)} = 1$ for all (t, x) implying $E|e^{itx}| = 1 < \infty$ for all t.
- Consider a symmetric density function, with $f(-x) = f(x)$ (symmetric around zero). Then resulting $\phi(t)$ is *real-valued*, and symmetric around zero.

- The CF completely determines the distribution of X (every cdf has a unique characteristic function).

- Let X have characteristic function $\phi_X(t)$. Then $Y = aX + b$ has characteristic function $\phi_Y(t) = e^{ibt}\phi_X(at)$.

- X and Y, independent, with characteristic functions $\phi_X(t)$ and $\phi_Y(t)$. Then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$

- $\phi(0) = 1$.

- For a given characteristic function $\phi_X(t)$ such that $\int_{-\infty}^{+\infty} |\phi_X(t)|dt < \infty$,[1] the corresponding density $f_X(x)$ is given by the inverse Fourier transform, which is

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \phi_X(t)\exp(-itx)dt.$$

Example: $N(0,1)$ distribution, with density $f(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$.

Take as given that the characteristic function of $N(0,1)$ is

$$\phi_{N(0,1)}(t) = \frac{1}{\sqrt{2\pi}}\int \exp\big(itx - x^2/2)\big)dx = \exp(-t^2/2). \tag{1}$$

Hence the inversion formula yields

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-t^2/2)\exp(-itx)dt.$$

Now making substitution $z = -t$, we get

$$\begin{aligned}&\frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\big(izx - z^2/2\big)dz\\ =&\frac{1}{\sqrt{2\pi}}\phi_{N(0,1)}(x) = \frac{1}{\sqrt{2\pi}}\exp(x^2/2) = f_{N(0,1)}(x). \quad \text{(Use Eq. (1))}\end{aligned}$$

- Characteristic function also summarizes the moments of a random variable. Specifically, note that the h-th derivative of $\phi(t)$ is

$$\phi^h(t) = \int_{-\infty}^{+\infty} i^h g(x)^h \exp(itg(x))f(x)dx. \tag{2}$$

[1] Here $|\cdot|$ denotes the modulus of a complex number. For $x + iy$, we have $|x + iy| = \sqrt{x^2 + y^2}$.

Hence, assuming the h-th moment, denoted $\mu^h_{g(x)} \equiv E[g(x)]^h$ exists, it is equal to

$$\mu^h_{g(x)} = \phi^h(0)/i^h.$$

Hence, assuming that the required moments exist, we can use Taylor's theorem to expand the characteristic function around $t = 0$ to get:

$$\phi(t) = 1 + \frac{it}{1}\mu^1_{g(x)} + \frac{(it)^2}{2}\mu^2_{g(x)} + ... + \frac{(it)^k}{k!}\mu^k_{g(x)} + o(t^k).$$

- **Cauchy distribution, cont'd:** The characteristic function for the Cauchy distribution is

  $$\phi(t) = \exp(-|t|).$$

  This is not differentiable at $t = 0$, which by Eq. (2) reflects the fact that its mean does not exist. Hence, the expansion of the characteristic function in this case is invalid.

  Moreover, consider the sample mean of iid Cauchy RV's: $\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ with each X_i distributed iid Cauchy. Then the c.f. of $\bar{X}_n$ is $\phi_n(t) = \mathbb{E}(\exp(it\bar{X}_n)) = \prod_{j=1}^n \mathbb{E}\exp(itX_j/n) = \prod_j \exp(-|t/n|) = \exp(-|t|)$. Averaging over multiple samples from the Cauchy distribution do not reduce the variance.

■■■

■■■

Multiple random variables

N-dimensional random vector (i.e., vector of random variables) is a function from the sample space Ω to $\mathcal{R}^N$ (N-dimensional Euclidean space).

Example: 2-coin toss. $\Omega = \{HH, HT, TH, TT\}$.

Consider the random vector $\vec{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, where $X_1 = \mathbf{1}$(at least one head), and $X_2 = \mathbf{1}$(at least one tail).

Ω	$\vec{X}$
HH	(1,0)
HT	(1,1)
TH	(1,1)
TT	(0,1)

Assuming coin is fair, we can also derive the joint probability distribution function for the random vector $\vec{X}$.

$\vec{X}$	$P_{\vec{X}}$
(1,0)	1/4
(1,1)	1/2
(0,1)	1/4

■■■

From the joint probabilities, can we obtain the individual probability distributions for X_1 and X_2 singly?

Yes, since (for example)

$$P(X_1 = 1) = P(X_1 = 1, X_2 = 0) + P(X_1 = 1, X_2 = 1) = 1/4 + 1/2 = 3/4$$

so that you obtain the *marginal probability* that $X_1 = x$ by summing the probabilities of all the outcomes in which $X_1 = x$.

■■■

From the joint probabilities, can we derive the *conditional probabilities* (i.e., if we fixed a value for X_2, what is the conditional distribution of X_1 given X_2)?

Yes:

$$P(X_1 = 0|X_2 = 0) = 0$$
$$P(X_1 = 1|X_2 = 0) = 1$$

and

$$P(X_1 = 0|X_2 = 1) = 1/3$$
$$P(X_1 = 1|X_2 = 1) = 2/3$$

&etc.

Namely: $P(X_1|X_2 = x) = P(X_1, x)/P(X_2 = x)$

Note: conditional probabilities tell you nothing about causality.

■■■

For this simple example of the 2-coin toss, we have derived the fundamental concepts: (i) joint probability; (ii) marginal probability; (iii) conditional probability.

More formally, for continuous random variables, we can define the analogous concepts.

Definition 4.1.10:
A function $f_{X_1,X_2}(x_1, x_2)$ from $\mathcal{R}^2$ to $\mathcal{R}$ is called a *joint probability density function* if, for

every $A \subset \mathcal{R}^2$:

$$P((X_1, X_2) \in A) = \underbrace{\int\int}_{A} f_{X_1,X_2}(x_1, x_2)dx_1dx_2.$$

The corresponding *marginal* density function are given by

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)dx_2 \\ f_{X_2}(x_2) &= \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)dx_1. \end{aligned}$$

As before, for the marginal density of X_1, you "sum over" all possible values of X_2, holding X_1 fixed.

The corresponding *conditional* density functions are

$$\begin{aligned} f_{X_1|X_2}(x_1|x_2) &= \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)} = \frac{f_{X_1,X_2}(x_1, x_2)}{\int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)dx_1} \\ f_{X_2|X_1}(x_2|x_1) &= \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)} = \frac{f_{X_1,X_2}(x_1, x_2)}{\int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)dx_2}. \end{aligned}$$

By rewriting the above as

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_2|X_1}(x_2|x_1)f_{X_1}(x_1)}{\int_{-\infty}^{\infty} f_{X_2|X_1}(x_2|x_1)f_{X_1}(x_1)dx_1}$$

we obtain *Baye's Rule* for multivariate random variables. In the Bayesian context, the above expression is interpreted as the "posterior density of x_1 given x_2".

These are all density functions: the joint, marginal and conditional density functions all integrate up to 1.

■■■

Independence of random variables

X_1 and X_2 are independent iff, for all (x_1, x_2),

$$\begin{aligned} P(X_1 \le x_1; X_2 \le x_2) &= F_{X_1,X_2}(x_1, x_2) \\ &= F_{X_1}(x_1) * F_{X_2}(x_2) = P(X_1 \le x_1) \cdot P(X_2 \le x_2) \end{aligned}$$

When the density exists, we can express independence also as, for all (x_1, x_2),

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) * f_{X_2}(x_2)$$

which implies

$$f_{X_1|X_2}(x_1|x_2) = f_{X_1}(x_1)$$
$$f_{X_2|X_1}(x_2|x_1) = f_{X_2}(x_2).$$

■■■

For conditional densities, it is natural to define:

Conditional expectation:

$$E(X_1|X_2 = x_2) = \int_{-\infty}^{\infty} x f_{X_1|X_2}(x|x_2)dx.$$

Conditional CDF:

$$F_{X_1|X_2}(x_1|x_2) = Prob(X_1 \le x_1|X_2 = x_2) = \int_{-\infty}^{x_1} f_{X_1|X_2}(x|x_2)dx.$$

Conditional CDF can be viewed as a special case of a conditional expectation: $E\left[\mathbf{1}(X_1 \le x_1)|X_2\right]$.

■■■

Example: X_1, X_2 distributed uniformly on the triangle $(0,0), (0,1), (1,0)$: that is,

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} 2 & \text{if } x_1 + x_2 \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Marginals:

$$f_{X_1}(x_1) = \int_0^{1-x_1} 2dx_2 = 2 - 2x_1$$
$$f_{X_2}(x_2) = \int_0^{1-x_2} 2dx_1 = 2 - 2x_2$$

Hence, $E(X_1) = \int_0^1 x_1(2-2x_1)dx_1 = 2\int_0^1 (x_1 - x_1^2)dx_1 = 2\left[\frac{1}{2}x_1^2 - \frac{1}{3}x_1^3\right]_0^1 = \frac{1}{3}$.

$Var(X_1) = EX_1^2 - (EX_1)^2 = \frac{1}{6} - (\frac{1}{3})^2 = \frac{1}{18}$

Note: $f_{X_1,X_2}(x_1,x_2) \neq f_{X_1}(x_1) * f_{X_2}(x_2)$: so not independent.

Conditionals:

$$f_{X_1|X_2}(x_1|x_2) = 2/(2-2x_2), \text{ for } 0 \le x_1 \le 1-x_2$$
$$f_{X_2|X_1}(x_2|x_1) = 2/(2-2x_1)$$

so

$$E(X_1|X_2) = \int_0^{1-x_2} x_1 \frac{2}{2-2x_2} dx_1 = \frac{2}{2-2x_2}\left[\frac{1}{2}x_1^2\right]_0^{1-x_2} = \frac{1-x_2}{2}.$$
$$E(X_1^2|X_2) = \int_0^{1-x_2} x_1^2 \frac{1}{1-x_2} dx_1 = \frac{1}{1-x_2}\left[\frac{1}{3}x_1^3\right]_0^{1-x_2} = \frac{1}{3} * (1-x_2)^2$$

so that

$$Var(X_1|X_2) = E(X_1^2|X_2) - [E(X_1|X_2)]^2 = \frac{1}{12}(1-x_2)^2.$$

■■■

Note: a useful way to obtain a marginal density is to use the conditional density formula:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)dx_2 = \int_{-\infty}^{\infty} f_{X_1|X_2}(x_1|x_2) f_{X_2}(x_2)dx_2.$$

This also provides an alternative way to calculate the marginal mean EX_1:

$$EX_1 = \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1)dx_1 = \int_{-\infty}^{\infty} x_1 \left[\int_{-\infty}^{\infty} f_{X_1|X_2}(x_1|x_2) f_{X_2}(x_2)dx_2\right] dx_1$$
$$\Rightarrow EX_1 = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x_1 f_{X_1|X_2}(x_1|x_2)dx_1\right] f_{X_2}(x_2)dx_2$$
$$= E_{X_2} E_{X_1|X_2} X_1$$

which is the *Law of iterated expectations.*

(In the last line of the above display, the subscripts on the expectations indicate the probability distribution that we take the expectations with respect to.)

■■■

Similar expression exists for variance:

$$Var X_1 = E_{X_2} Var_{X_1|X_2}(X_1) + Var_{X_2} E_{X_1|X_2}(X_1).$$

■■■

- Truncated random variables: Let (X, Y) be jointly distributed according to the joint density function $f_{X,Y}$, with support $\mathcal{X} \times \mathcal{Y}$.

  Then the random variables *truncated* to the region $A \in \mathcal{X} \times \mathcal{Y}$ follow the density

$$\frac{f_{X,Y}(x,y)}{Prob_{X,Y}\left(X, Y \in A\right)} = \frac{f_{X,Y}(x,y)}{\int \int_A f_{X,Y}(x,y) dx dy}$$

  with support $(X, Y) \in A$.

- **Multivariate characteristic function**

  Let $\vec{X} \equiv (X_1, \ldots, X_m)'$ denote an m-vector of random variables with joint density $f_{\vec{X}}(\vec{x})$.

$$\begin{aligned} \phi_{\vec{X}}(t) &= \mathbb{E} \exp(it'\vec{x}) \\ &= \int_{-\infty}^{+\infty} \exp(it'\vec{x}) f_{\vec{X}}(\vec{x}) d\vec{x} \end{aligned} \tag{3}$$

  where t is an m-dimensional real vector.

  This suggests that any multivariate distribution is determined by the behavior of *linear combinations* of its components. **Cramer-Wold device**: a Borel probability measure on $\mathbb{R}^m$ is uniquely determined by the totality of its one-dimensional projections. (A formal statement will come later.)

  Clearly: $\phi(0, 0, \ldots, 0) = 1$

Transformations of multivariate random variables: some cases

1. $X_1, X_2 \sim f_{X_1,X_2}(x_1, x_2)$

Consider the random variable $Z = g(X_1, X_2)$.

CDF: $F_Z(z)$=Prob$(g(X_1, X_2) \leq z) = \int\int_{g(x_1,x_2)\leq z} f_{X_1,X_2}(x_1, x_2)dx_1dx_2$.

PDF: $f_Z(z) = \frac{\partial F_Z(z)}{\partial z}$.

Example: triangle problem again; consider $g(X_1, X_2) = X_1 + X_2$.

First, note that support of Z is $[0, 1]$.

$$\begin{aligned}
F_Z(z) &= Prob(X_1 + X_2 \leq z) \\
&= \int_0^z \int_0^{z-x_1} 2dx_2dx_1 \\
&= 2\int_0^z (z - x_1)dx_1 \\
&= 2(z^2 - \frac{1}{2}z^2) = z^2.
\end{aligned}$$

Hence, $f_z(z) = 2z$.

■■■

2. Convolution: $X \sim f_X$, $e \sim f_e$, with (X, e) independent. Let $Y = X + e$. What is f_y?

(Ex: measurement error. Y is contaminated version of X)

$$\begin{aligned}
F_y(y) = P(X + e < y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{y-e} f_X(x)f_e(e)dxde \\
&= \int_{-\infty}^{+\infty} F_X(y - e)f_e(e)de \\
\Rightarrow \quad f_y(y) &= \int_{-\infty}^{+\infty} f_X(y - e)f_e(e)de \\
&= \int_{-\infty}^{+\infty} f_X(x)f_e(y - x)dx.
\end{aligned}$$

Recall: $\phi_Y(t) = \phi_X(t)\phi_e(t) \;\Rightarrow\; \phi_X(t) = \frac{\phi_Y(t)}{\phi_e(t)}$. This is "deconvolution".

$\blacksquare\blacksquare\blacksquare$

3. Bivariate change of variables

$X_1, X_2 \sim f_{X_1,X_2}(x_1, x_2)$

$Y_1 = g_1(X_1, X_2)$, $Y_2 = g_2(X_1, X_2)$. What is joint density $f_{Y_1,Y_2}(y_1, y_2)$?

CDF:

$$\begin{aligned} F_{Y_1,Y_2}(y_1, y_2) &= Prob(g_1(X_1, X_2) \leq y_1, g_2(X_1, X_2) \leq y_2) \\ &= \int\int_{g_1(x_1,x_2)\leq y_1,\ g_2(x_1,x_2)\leq y_2} f_{X_1,X_2}(x_1, x_2) dx_1 dx_2. \end{aligned}$$

PDF: assume that the mapping from (X_1, X_2) to (Y_1, Y_2) is one-to-one, which implies that the system $\left\{ \begin{array}{c} y_1 = g_1(x_1, x_2) \\ y_2 = g_2(x_1, x_2) \end{array} \right\}$ can be inverted to get $\left\{ \begin{array}{c} x_1 = h_1(y_1, y_2) \\ x_2 = h_2(y_1, y_2) \end{array} \right\}$.

Define the Jacobian matrix $J_h = \begin{bmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} \end{bmatrix}$.

Then the bivariate change of variables formula is:

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}\left(h_1(y_1, y_2), h_2(y_1, y_2)\right) * |J|$$

where $|J_h|$ denotes the absolute value of the determinant of J_h.

(see heuristic proof in the appendix)

$\blacksquare\blacksquare\blacksquare$

$\blacksquare\blacksquare\blacksquare$

Example: Triangle problem again

Consider

$$\begin{aligned} Y_1 &= g_1(X_1, X_2) = X_1 + X_2 \\ Y_2 &= g_2(X_1, X_2) = X_1 - X_2 \end{aligned} \tag{4}$$

Jacobian matrix: inverse mappings are

$$\begin{aligned} X_1 &= \frac{1}{2}(Y_1 + Y_2) \equiv h_1(Y_1, Y_2) \\ X_2 &= \frac{1}{2}(Y_1 - Y_2) \equiv h_2(Y_1, Y_2) \end{aligned} \tag{5}$$

so $J = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$ and $|J| = \frac{1}{2}$.

Hence,

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{2} \cdot f_{X_1,X_2}(\frac{1}{2}(y_1 + y_2), \frac{1}{2}(y_1 - y_2)) = 1,$$

a uniform distribution.

Support of (Y_1, Y_2):
(i) From Eqs. (4), you know $Y_1 \in [0, 1]$, $Y_2 \in [-1, 1]$
(ii) $0 \leq X_1 + X_2 \leq 1 \Rightarrow 0 \leq Y_1 \leq 1$. Redundant.
(iii) $0 \leq X_1 \leq 1 \Rightarrow 0 \leq \frac{1}{2}(Y_1 + Y_2) \leq 1$. Only lower inequality is new, so $Y_1 \geq -Y_2$
(iv) $0 \leq X_2 \leq 1 \Rightarrow 0 \leq \frac{1}{2}(Y_1 - Y_2) \leq 1$. Only lower inequality is new, so $Y_1 \geq Y_2$.

Graph:

■■■

Covariance and Correlation

Notation: $\mu_1 = EX_1$, $\mu_2 = EX_2$, $\sigma_1^2 = VarX_1$, $\sigma_2^2 = VarX_2$.

Covariance:

$$\begin{aligned} Cov(X_1, X_2) &= E\left[(X_1 - \mu_1) \cdot (X_2 - \mu_2)\right] \\ &= E(X_1 X_2) - \mu_1 \mu_2 \\ &= E(X_1 X_2) - EX_1 EX_2 \end{aligned}$$

taking values in $(-\infty, \infty)$. (Obviously, $Cov(X, X) = Var(X)$.)

Correlation:

$$Corr(X_1, X_2) \equiv \rho_{X_1,X_2} = \frac{Cov(X_1, X_2)}{\sigma_1 \sigma_2}$$

which is bounded between $[-1, 1]$.

■■■

Example: triangle problem again

Earlier, we showed $\mu_1 = \mu_2 = 1/3$ and $\sigma_1^2 = \sigma_2^2 = \frac{1}{18}$.

$EX_1X_2 = 2\int_0^1 \int_0^{1-x_1} x_1x_2dx_2dx_1 = 1/12$

Hence

$$\begin{aligned} Cov(X_1, X_2) &= \frac{1}{12} - (\frac{1}{3})^2 = -1/36 \\ Corr(X_1, X_2) &= \frac{-1/36}{1/18} = -1/2. \end{aligned}$$

■■■

Useful results:

- $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$. As we remarked before, Variance is not a linear operator.
- More generally, for $Y = \sum_{i=1}^n X_i$, we have
$$Var(Y) = \sum_{i=1}^{n} Var(X_i) + \sum_{i<j} 2\ Cov(X_i, X_j).$$
- If X_1 and X_2 are independent, then $Cov(X_1, X_2) = 0$. Important: the converse is not true: zero covariance does not imply independence. Covariance only measures (roughly) a linear relationship between X_1 and X_2.

  Example: $X \sim U[-1, 1]$ and consider $Cov(X, X^2)$

■■■

(Example: Auctions and the Winner's Curse) Two bidders participate in an auction for a painting. Each bidder has the *same* underlying valuation for the painting, given by the random variable $V \sim U[0, 1]$. Neither bidder knows V.

Each bidder receives a signal about V: $X_i|V \sim U[0, V]$, and X_1 and X_2 are independent, conditional on V (i.e., $f_{X_1,X_2}(x_1, x_2|V) = f_{X_1}(x_1|V) \cdot f_{X_2}(x_2|V)$).

(a) Assume each bidder submits a bid equal to her conditional expectation: for bidder 1, this is $E(V|X_1)$. How much does she bid?

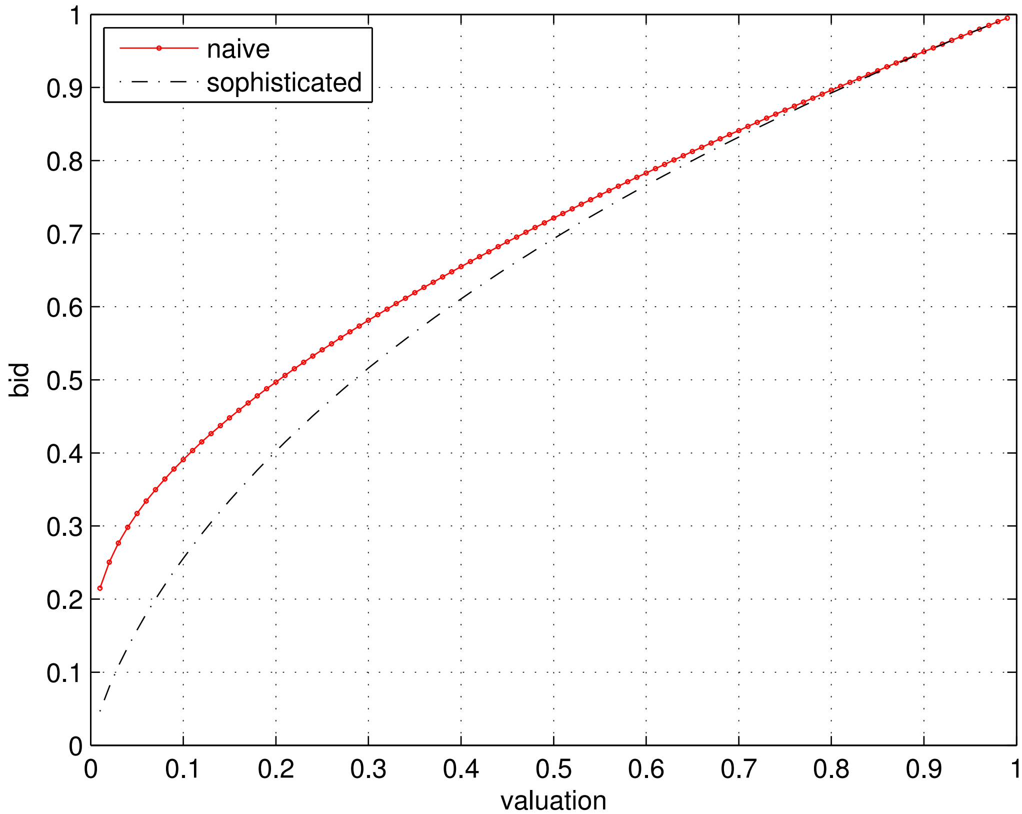
(b) Given this way of bidding, bidder 1 wins if and only if $X_1 > X_2$: that is, her signal is higher than bidder 2's signal. What is bidder 1's expected revenue conditional on winning, that is, her conditional expectation of the value V, given both her signal X_1 and the event that she wins: that is, $E\left[V|X_1, X_1 > X_2\right]$?

Solution (use Baye's Rule in both steps):

- Part (a):
  - $f(v|x_1) = \frac{f(x_1|v)f(v)}{\int_{x_1}^1 f(x_1|v)f(v)dv} = \frac{1/v}{\int_{x_1}^1 1/vdv} = -1/(v\log x_1)$.
  - Hence: $E[v|x_1] = \frac{-1}{\log x_1}\int_{x_1}^1 (v/v)dv = \frac{-1}{\log x_1}(1-x_1) = \frac{(1-x_1)}{-\log x_1}$.
- Part (b):

$$E(v|x_1, x_2 < x_1) = \int vf(v|x_1, x_2 < x_1)dv = \frac{\int vf(x_1, v|x_2 < x_1)dv}{\int f(x_1, v|x_2 < x_1)dv}$$

  - $f(v, x_1, x_2) = f(x_1, x_2|v) \cdot f(v) = 1/v^2$.
  - $Prob(x_2 < x_1|v) = \int_0^v \int_0^{x_1} \frac{1}{v^2} dx_2 dx_1 = \frac{1}{v^2}\int_0^v x_2 dx_1 = 1/2$. Hence also unconditional $Prob(x_2 < x_1) = 1/2$.
  - $f(v, x_1, x_2|x_1 > x_2) = \frac{f(v,x_1,x_2)}{P(x_1>x_2)} = 2/v^2$.
  - $f(v, x_1|x_1 > x_2) = \int_0^{x_1} f(v, x_1, x_2|x_1 > x_2)dx_2 = \frac{2x_1}{v^2}$
  - $E(v|x_1, x_2 > x_2) = \frac{\int_{x_1}^1 v\ f(v,x_1|x_1>x_2)dv}{\int_{x_1}^1 f(v,x_1|x_1>x_2)dv} = \frac{\int_{x_1}^1 \frac{2x_1}{v}dv}{\int_{x_1}^v \frac{2x_1}{v^2}dv} = \frac{-2x_1\log x_1}{-2x_1(1-1/x_1)}$
  - Hence: posterior mean is $\frac{-x_1\log x_1}{1-x_1}$.
  - Graph results of part (a) vs. part (b). The feature that the line for part (b) lies below that for part (a) is called the "winner's curse": if bidders bid naively (i.e., according to (a)), their expectated profit is negative.



Discussion.

- Example of *adverse selection*: event of winning selects most overly optimistic bidder.
- In equilibrium: bidders will bid more cautiously to avoid winner's curse.[2]
- More generally: "pivotal event" (event that your action affects your payoffs) conveys information which counteracts your own private information
  - Pivotal jury voting: with unanimity rule, your vote is "pivotal" (makes a difference) only when everyone else on the jury has voted to convict.
  - Market microstructure: other traders' willingness-to-trade counteracts your desire to trade.

[2] Milgrom and Weber (1982, *Econometrica*)

- All these examples assume that agent's valuations are positively related. What if they were negatively related?
  - Auction for painting which may or may not be by Rembrandt.
  - Rembrandt lover vs. Rembrandt hater.
  - Each bidders receives noisy signal of whether painting is by Rembrandt.
  - Does winner's curse arise?

A Appendix: additional material

A.1 Moment generating function

[SKIP] Moment generating function

The moments of a random variable are summarized in the *moment generating function*.

Definition: the moment-generating function of X is $M_X(t) \equiv E\exp(tX)$, provided that the expectation exists in some neighborhood $t \in [-h, h]$ of zero.

Specifically:

$$M_x(t) = \begin{cases} \int_{-\infty}^{\infty} e^{tx} f_X(x)dx & \text{for } X \text{ continuous} \\ \sum_{x \in \mathcal{X}} e^{tx} P(X = x) & \text{for } X \text{ discrete.} \end{cases}$$

Series expansion around $t = 0$: Note that

$$M_X(t) = Ee^{tX} = 1 + tEX + \frac{t^2}{2}EX^2 + \frac{t^3}{6}EX^3 + ... + \frac{t^n}{n!}EX^n + ...$$

so that the uncentered moments of X are generated from this function by:

$$EX^n = M_X^{(n)}(0) \equiv \frac{d^n}{dt^n} M_X(t)\bigg|_{t=0},$$

which is the n-th derivative of the MGF, evaluated at $t = 0$.

When it exists (see below), then MGF provides alternative description of a probability distribution. Mathematically, it is a *Laplace transform*.

Example: standard normal distribution:

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(tx - \frac{x^2}{2}\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}((x-t)^2 - t^2)\right) dx \\
&= \exp(\frac{1}{2}t^2) \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-t)^2\right) dx \\
&= \exp(\frac{1}{2}t^2) \cdot 1
\end{aligned}$$

where last term on RHS is integral over density function of $N(t,1)$, which integrates to one.

First moment: $EX = M_X^1(0) = t \cdot \exp(\frac{1}{2}t^2)\big|_{t=0} = 0$.
Second moment: $EX^2 = M_X^2(0) = \exp(\frac{1}{2}t^2) + t^2 \exp(\frac{1}{2}t^2) = 1$.

■■■

In many cases, the MGF can characterize a distribution. But problem is that it may not exist (eg. Cauchy distribution)

For a RV X, is its distribution uniquely determined by its moment generating function?

Thm 2.3.11: For $X \sim F_X$ and $Y \sim F_Y$, if M_X and M_Y exist, and $M_X(t) = M_Y(t)$ for all t in some neighborhood of zero, then $F_X(u) = F_Y(u)$ for all u.

If the MGF exists, then it characterizes a random variable with an *infinite* number of moments (because the MGF is infinitely differentiable).

Ex: log-normality. If $X \sim N(0,1)$, then $Y = \exp(X)$ is log-normal distributed. We have $EY = e^{1/2}$ and generally $EY^m = e^{m^2/2}$ so all the moments exist. But $E\exp(tY) \to \infty$ for all t. Note that by the expansion (for t around zero)

$$\begin{aligned}
Ee^{tY} &= 1 + te^{1/2} + \frac{t^2}{2}e^1 + ... + \frac{t^n}{n!}e^{n/2} + ... \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \frac{t^i}{i!} e^{i/2} \to \infty.
\end{aligned}$$

.1 Heuristic argument for bivariate change of variables

[skip] To get some intuition for the above result, consider the probability that the random variables (y_1, y_2) lie within the rectangle

$$\left\{ \underbrace{(y_1^*, y_2^*)}_{\equiv A}, \underbrace{(y_1^* + dy_1, y_2^*)}_{\equiv B}, \underbrace{(y_1^*, y_2^* + dy_2)}_{\equiv C}, \underbrace{(y_1^* + dy_1, y_2^* + dy_2)}_{\equiv D} \right\}.$$

For $dy_1 > 0, dy_2 > 0$ small, this is approximately

$$f_{y_1,y_2}(y_1^*, y_2^*)dy_1dy_2 \tag{6}$$

which, in turn, is approximately

$$f_{x_1,x_2}(\underbrace{h_1(y_1^*, y_2^*)}_{\equiv h_1^*}, \underbrace{h_2(y_1^*, y_2^*)}_{\equiv h_2^*})\text{“}dx_1dx_2\text{”}. \tag{7}$$

In the above, dx_1 is the change in x_1 occasioned by the changes from y_1^* to $y_1^* + dy_1$ and from y_2^* to $y_2^* + dy_2$.

Eq. (6) is the area of the rectangle formed from points (A, B, C, D) multiplied by the density $f_{y_1,y_2}(y_1^*, y_2^*)$. Similarly, Eq. (7) is the density $f_{x_1,x_2}(h_1^*, h_2^*)$ multiplying "dx_1dx_2", which is the area of a parallelogram defined by the four points (A', B', C', D'):

$$\begin{aligned}
A = (y_1^*, y_2^*) &\to A' = (h_1^*, h_2^*) \\
B = (y_1^* + dy_1, y_2^*) &\to B' = (h_1(B), h_2(B)) \approx (h_1^* + dy_1\frac{\partial h_1}{\partial y_1^*}, h_2^* + dy_1\frac{\partial h_2}{\partial y_1^*}) \\
C = (y_1^*, y_2^* + dy_2) &\to C' \approx (h_1^* + dy_2\frac{\partial h_1}{\partial y_2^*}, h_2^* + dy_2\frac{\partial h_2}{\partial y_2^*}) \\
D = (y_1^* + dy_1, y_2^* + dy_2) &\to D' \approx (h_1^* + dy_1\frac{\partial h_1}{\partial y_1^*} + dy_2\frac{\partial h_1}{\partial y_2^*}, h_2^* + dy_1\frac{\partial h_2}{\partial y_1^*} + dy_2\frac{\partial h_2}{\partial y_2^*})
\end{aligned}$$

In defining the points B', C', D', we have used first-order approximations of $h_1(y_1^*, y_2^* + dy_2)$ around (y_1^*, y_2^*); etc.

The area of $(A'B'C'D')$ is the same as the area of the parallelogram formed by the two vectors

$$\vec{a} \equiv \left(dy_1\frac{\partial h_1}{\partial y_1^*}, dy_1\frac{\partial h_2}{\partial y_1^*}\right)'; \quad \vec{b} \equiv \left(dy_2\frac{\partial h_1}{\partial y_2^*}, dy_2\frac{\partial h_2}{\partial y_2^*}\right)'.$$

The area of this is given by the length of the cross-product

$$|\vec{a} \times \vec{b}| = |\det\ [\vec{a}, \vec{b}]| = dy_1dy_2\left|\frac{\partial h_1}{\partial y_1^*}\frac{\partial h_2}{\partial y_2^*} - \frac{\partial h_1}{\partial y_2^*}\frac{\partial h_2}{\partial y_1^*}\right| = dy_1dy_2|J_h|.$$

Hence, by equating (6) and (7) and substituting in the above, we obtain the desired formula.