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

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

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

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

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.

■■■

Multivariate CDFs

Consider two random variables (x_1, x_2). The bivariate CDF F_{x_1,x_2} is defined as

$$F_{x_1,x_2}(a,b) = Pr(x_1 \leq a, x_2 \leq b).$$

When (x_1, x_2) have a joint density function, then the joint CDF equals

$$F_{x_1,x_2}(a,b) = \int_{x_1:-\infty}^{a} \int_{x_2:-\infty}^{b} f_{x_1,x_2}(x_1,x_2)dx_2dx_1.$$

Properties of F_{x_1,x_2}:

1. $\lim_{x_j \to -\infty} F(x_1, x_2) = 0, \; j = 1, 2.$

2. $\lim_{x_1 \to +\infty, x_2 \to +\infty} F(x_1, x_2) = 1.$

3. (rectangle inequality): for all $(a_1, a_2), (b_1, b_2)$ such that $a_1 < b_1$, $a_2 < b_2$,

$$F(b_1, b_2) - F(a_1, b_2) - [F(b_1, a_2) - F(a_1, a_2)] \geq 0.$$

   When F has second-order derivatives, this is equivalent to $\frac{\partial^2 F}{\partial x_1 \partial x_2} \geq 0$ (supermodularity).

4. Marginalization: $F_{x_1,x_2}(a, \infty) = F_{x_1}(a)$ (marginal CDF of x_1). Similarly for F_{x_2}.

5. $F_{x_1,x_2}(\cdot, \cdot)$ is increasing in both arguments.

These properties can be generalized straightforwardly to the n-variate CDF $F_{x_1,\ldots,x_n}$. For this case, property 3 above becomes:

- (rectangle inequality, n-variate): for all $(a_1, \ldots, a_n)$, $(b_1, \ldots, b_n)$ with $a_i < b_i$ for $i = 1, \ldots, n$

$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\cdots+i_n} F(x_{1,i_1}, x_{2,i_2}, \ldots, x_{n,i_n}) \geq 0$$

where, for all $j = 1, \ldots, n$, we have $x_{j,1} = a_j$, $a_{j,2} = b_j$. When F has n-variate derivatives, then the condition becomes

$$\frac{\partial^n F}{\partial x_1 \partial x_2, \ldots, \partial x_n} \geq 0.$$

■■■

Independence of random variables

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

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

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

$$\begin{aligned} f_{X_1|X_2}(x_1|x_2) &= f_{X_1}(x_1) \\ f_{X_2|X_1}(x_2|x_1) &= f_{X_2}(x_2). \end{aligned}$$

■■■

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:

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

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

■■■

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

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

Marginals:

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

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:

$$\begin{aligned} &f_{X_1|X_2}(x_1|x_2) = 2/(2-2x_2), \text{ for } 0 \leq x_1 \leq 1 - x_2 \\ &f_{X_2|X_1}(x_2|x_1) = 2/(2-2x_1) \end{aligned}$$

so

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

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:

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

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:

$$VarX_1 = E_{X_2} Var_{X_1|X_2}(X_1) + Var_{X_2} E_{X_1|X_2}(X_1).$$

■■■

- 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

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

  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 $f_{\vec{X}}(\vec{x})$.

  $$
  \begin{aligned}
  \phi_{\vec{X}}(t) &= \mathbb{E} \exp(it'\vec{x}) \\
  &= \int_{-\infty}^{+\infty} \exp(it'\vec{x}) f_{\vec{X}}(\vec{x}) d\vec{x}
  \end{aligned}
  \tag{1}
  $$

  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.

  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".

■■■

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:

$$\begin{aligned} F_{Y_1,Y_2}(y_1, y_2) &= Prob(g_1(X_1, X_2) \leq y_1, g_2(X_1, X_2) \leq y_2) \\ &= \int\int_{g_1(x_1,x_2)\leq y_1,\ g_2(x_1,x_2)\leq y_2} f_{X_1,X_2}(x_1, x_2)dx_1dx_2. \end{aligned}$$

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.

■■■

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

$$f_{y_1,y_2}(y_1^*, y_2^*)dy_1dy_2 \tag{2}$$

which, in turn, is approximately

$$f_{x_1,x_2}(\underbrace{h_1(y_1^*, y_2^*)}_{\equiv h_1^*}, \underbrace{h_2(y_1^*, y_2^*)}_{\equiv h_2^*})\text{“}dx_1dx_2\text{”}. \tag{3}$$

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. (2) 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. (3) is the density $f_{x_1,x_2}(h_1^*, h_2^*)$ multiplying "$dx_1 dx_2$", which is the area of a parallelogram defined by the four points (A', B', C', D'):

$$\begin{aligned} A = (y_1^*, y_2^*) &\rightarrow A' = (h_1^*, h_2^*) \\ B = (y_1^* + dy_1, y_2^*) &\rightarrow B' = (h_1(B), h_2(B)) \approx (h_1^* + dy_1 \frac{\partial h_1}{\partial y_1^*}, h_2^* + dy_1 \frac{\partial h_2}{\partial y_1^*}) \\ C = (y_1^*, y_2^* + dy_2) &\rightarrow C' \approx (h_1^* + dy_2 \frac{\partial h_1}{\partial y_2^*}, h_2^* + dy_2 \frac{\partial h_2}{\partial y_2^*}) \\ D = (y_1^* + dy_1, y_2^* + dy_2) &\rightarrow D' \approx (h_1^* + dy_1 \frac{\partial h_1}{\partial y_1^*} + dy_2 \frac{\partial h_1}{\partial y_2^*}, h_2^* + dy_1 \frac{\partial h_2}{\partial y_1^*} + dy_2 \frac{\partial h_2}{\partial y_2^*}) \end{aligned} \tag{4}$$

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

$$|\vec{a} \times \vec{b}| = |\det\ [\vec{a}, \vec{b}]| = dy_1 dy_2 \left| \frac{\partial h_1}{\partial y_1^*} \frac{\partial h_2}{\partial y_2^*} - \frac{\partial h_1}{\partial y_2^*} \frac{\partial h_2}{\partial y_1^*} \right| = dy_1 dy_2 |J_h|.$$

Hence, by equating (2) and (3) and substituting in the above, we obtain the desired formula.

■■■

Example: Triangle problem again

Consider

$$\begin{aligned} Y_1 &= g_1(X_1, X_2) = X_1 + X_2 \\ Y_2 &= g_2(X_1, X_2) = X_1 - X_2 \end{aligned} \tag{5}$$

Jacobian matrix: inverse mappings are

$$\begin{aligned} X_1 &= \frac{1}{2}(Y_1 + Y_2) \equiv h_1(Y_1, Y_2) \\ X_2 &= \frac{1}{2}(Y_1 - Y_2) \equiv h_2(Y_1, Y_2) \end{aligned} \tag{6}$$

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. (5), 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 $Cov(X, X^2)$.

■■■

Practice: assume $X, Y \sim U[0, 1]^2$ (distributed uniformly on the unit square; $f(x, y) = 1$)

What is:

1. $f(X, Y|Y = \frac{1}{2})$
2. $f(X, Y|Y \geq \frac{1}{2})$
3. $f(X|Y = \frac{1}{2})$

4. $f(X|Y)$
5. $f(Y|Y \geq \frac{1}{2})$
6. $f(X|Y \geq \frac{1}{2})$
7. $f(X, Y|Y \geq X)$
8. $f(X|Y \geq X)$

(4) is $\frac{f(x,y)}{f(y)}$ where $f(y) = \int_0^1 f(x, y)dx = 1$.

From (4), (3) is special case, and (1) is equivalent to (3).

(2) is $\frac{f(x,y)}{Prob(y \geq \frac{1}{2})} = 2f(x, y) = 2$. Then obtain (5) and (6) by integrating this density over the appropriate range.

(7) is $\frac{f(x,y)}{Prob(y \geq x)} = 1/\frac{1}{2} = 2$, over the region $0 \leq X \geq 1;\ Y \geq X$. Then (8) is the marginal of this: $f(x|y \geq x) = \int_x^1 2dy = 2(1 - x)$.

Two additional problems:

1. **(Sample selection bias)** Let X denote number of children, and Y denote years of schooling. We make the following assumptions:

   - X takes values in $[0, 2\theta]$, where $\theta > 0$. θ is unknown.
   - Y is renormalized to take values in $[0, 1]$, with $Y = \frac{1}{2}$ denoting completion of high school.
   - In the population, (X, Y) are jointly uniformly distributed on the triangle

   $$\left\{(x, y) : x \in [0, 2\theta],\ y \leq 1 - \frac{1}{2\theta}x\right\}.$$

   Suppose you know that the average number of children among high school graduates is 2. What is the average number of children in the population?

   Solution: Use the information that $E[X|Y > \frac{1}{2}] = 2$ to recover the value of θ.

   - Joint density of (X, Y) is $\frac{1}{\theta}$ on this triangle.
   - $P(Y \geq \frac{1}{2}) = \int_{1/2}^{1} \int_{0}^{2\theta(1-y)} \frac{1}{\theta} dy = \int_{1/2}^{1} 2(1-y) dy = 2(y - \frac{1}{2}y^2) = \frac{1}{4}$.
   - Marginal $f(X) = \frac{1}{\theta} \int_0^{1-\frac{1}{2\theta}X} dy = \frac{1}{\theta}\left(1 - \frac{1}{2\theta}X\right)$. So that $EX = \frac{2}{3}\theta$.
   - Define $g(X, Y) \equiv f(X, Y|Y \geq 1/2) = \frac{f(X,Y)}{P(Y \geq 1/2)} = \frac{4}{\theta}$, on the triangle $X \in [0, \theta],\ Y \in [1/2, 1],\ Y \leq 1 - \frac{1}{2\theta}X$.
   - Marginal $g(X) = \int_{1/2}^{1-\frac{1}{2\theta}X} \frac{4}{\theta} dy = \frac{2}{\theta}\left(1 - \frac{X}{\theta}\right)$.
   - $E(X|Y \geq 1/2) = \int_0^\theta X g(X) dX = \frac{2}{\theta} \int_0^\theta X - \frac{1}{\theta}X^2 dX = \frac{\theta}{3}$.
   - Therefore if $E(X|Y \geq 1/2) = 2$ then $\theta = 6$, and $EX = \frac{2}{3}\theta = 4$.

   Are there alternative ways to solve? Can use Baye's Rule $f(X|Y \geq 1/2) = \frac{P(Y \geq 1/2|X) f(X)}{\int_0^\theta P(Y \geq 1/2|X) f(X)}$, but this is not any easier (still need to derive the conditional $f(Y|X)$ and the marginal $f(X)$).

2. **(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 $E\left(V|X_1\right)$. How much does she bid?

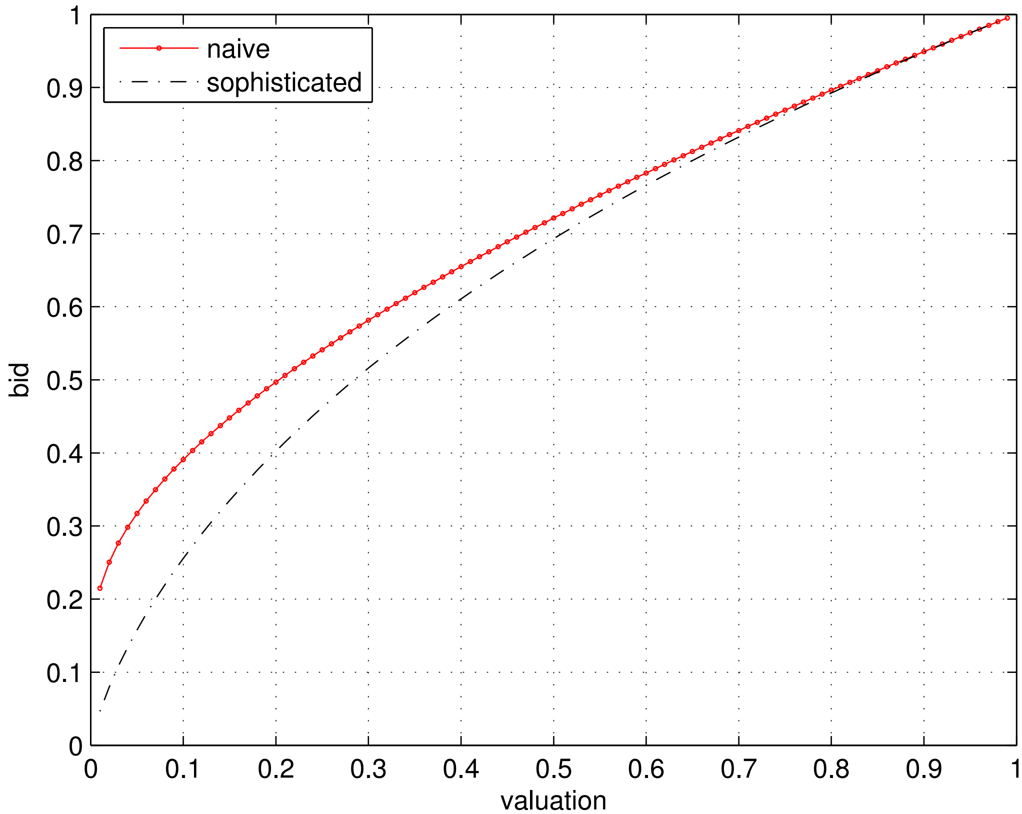
(b) Note that given this way of bidding, bidder 1 wins if and only if $X_1 > X_2$: that is, her signal is higher than bidder 2's signal. What is bidder 1's conditional expectation of the value V, given both her signal X_1 and the event that she wins: that is, $E\left[V|X_1, X_1 > X_2\right]$?

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.


naive
sophisticated
bid
valuation
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1