Introduction to probability

Lectures prepared by:
Elchanan Mossel
Yelena Shvets

Follows Jim Pitman’s book:
Probability
Sections 6.3
Conditional density

Example: \((X,Y)\) uniform in the unit disk centered at \(0\).

Question: \(P(X > 0.1 | Y > 0.1) = ?\)

Answer:
\[
P(X > 0.1 | Y > 0.1) = \frac{P(X > 0.1 \& Y > 0.1)}{P(Y > 0.1)}
\]

Question: \(P(X > 0.1 | Y = 0.1) = ?\)

Try:
\[
P(X > 0.1 | Y = 0.1) = \frac{P(X > 0.1 \& Y = 0.1)}{P(Y = 0.1)} = \frac{0}{0}!
\]
Infinitesimal Conditioning Formula

Solution: Replace by infinitesimal ratio:

\[
P[A|X = x] = \lim_{dx \to 0} \frac{P[A, X \in (x, x + dx)]}{P[X \in (x, x + dx)]}
\]
Infinitesimal Conditioning Formula

\[ P(A \mid X = x) = \lim_{{\Delta x \to 0}} \frac{P(A \mid X \in \Delta x)}{P(X \in \Delta x)} \]

\[ = \lim_{{\Delta x \to 0}} \frac{P(A \& X \in \Delta x)}{P(X \in \Delta x)} \]
**Conditional Density**

**Claim:** Suppose \((X,Y)\) have a joint density \(f(x,y)\), and \(x\) is such that \(f_X(x) > 0\), the conditional density of \(Y\) given \(X=x\) is a probability density defined by

\[
f_Y(y \mid X = x) = \frac{f(x, y)}{f_X(x)}
\]

**“Proof”:**

\[
P\left(a \leq Y \leq b \mid X = x\right) = \lim_{x \to \infty} P\left(a \leq Y \leq b \mid X \in (x, x+dx)\right) = \lim_{x \to \infty} \frac{P\left(a \leq Y \leq b, X \in (x, x+dx)\right)}{f_X(x) \, dx} = \lim_{x \to \infty} \int_a^b f(x,y) \, dy \, dx / f_X(x) \, dx = \int_a^b f_Y(y \mid X = x) \, dy
\]
Joint density of \((X,Y)\): \(f(x,y)\);

Renormalized slice for given \(y = \text{conditional density } P[X \mid Y=y]\),
The area is of each slice is 1.

Slices through density for fixed \(y\)'s.

Area of slice at \(y\) = height of marginal distribution at \(y\).
## Conditioning Formulae

<table>
<thead>
<tr>
<th>Discrete</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Multiplication Rule:</td>
<td></td>
</tr>
<tr>
<td>[ P(x, y) = P(Y = y</td>
<td>X = x)P(X = x) ]</td>
</tr>
<tr>
<td>• Division Rule:</td>
<td></td>
</tr>
<tr>
<td>[ P(Y = y</td>
<td>X = x) = \frac{P(X = x, Y = y)}{P(X = x)} ]</td>
</tr>
<tr>
<td>• Bayes’ Rule:</td>
<td></td>
</tr>
<tr>
<td>[ P(X = x</td>
<td>Y = y) = \frac{P(Y = y</td>
</tr>
</tbody>
</table>
Example: Uniform on a triangle.

Suppose that a point \((X,Y)\) is chosen uniformly at random from the triangle \(A: \{(x,y): 0 \leq x, 0 \leq y, x+y \leq 2\}\).

Question: Find \(P(Y \geq 1 \mid X = x)\).

Solution 1: Since Area\((A) = 2\), for all \(S\), \(P(S) = \text{Area}(S)/2\).

So for \(0 \leq x < 1\):

\[
P(X \in \Delta x) = \frac{1}{2} \Delta x \ (2-x - \frac{1}{2} \Delta x);
\]

\[
P(Y > 1, X \in \Delta x) = \frac{1}{2} \Delta x \ (1-x - \frac{1}{2} \Delta x);
\]

\[
P(Y \geq 1 \mid X \in \Delta x) = P(Y > 1, X \in \Delta x) / P(X \in \Delta x)
= (1-x - \frac{1}{2} \Delta x)/(2-x - \frac{1}{2} \Delta x);
\]

\[\rightarrow (1-x)/(2-x) \text{ as } \Delta x \rightarrow 0.\]

So \(P(Y \geq 1 \mid X = x) = (1-x)/(2-x)\)

And for \(x \geq 1\): \(P(Y \geq 1 \mid X= x) = 0\).
Example: Uniform on a triangle.

Suppose that a point \((X,Y)\) is chosen uniformly at random from the triangle \(A: \{(x,y): 0 \leq x, 0 \leq y, x + y \leq 2\}\).

**Question:** Find \(P(Y \geq 1 \mid X = x)\).

**Solution 2.** using the conditional density \(f_Y(y \mid X=x)\).

For the uniform distribution on a triangle of area 2,

\[
f(x,y) = \begin{cases} 
1/2, & \text{if } 0 \leq x, 0 \leq y, x + y \leq 2; \\
0, & \text{otherwise.}
\end{cases}
\]

So for \(0 \leq x \leq 2\),

\[
f_X(x) = \int_0^\infty f(x,y) \, dy = \int_0^{2-x} \frac{1}{2} \, dy = \frac{2-x}{2},
\]

\[
f_Y(y \mid X = x) = \begin{cases} 
\frac{f(x,y)}{f_X(x)} = \frac{1}{2-x}, & \text{if } 0 \leq y \leq 2 - x; \\
0, & \text{otherwise.}
\end{cases}
\]

Given, \(X=x\), the distribution of \(Y\) is uniform on \((0,2-x)\).

\[
P[Y \geq 1 \mid X = x] = \int_1^{2-x} \frac{1}{2-x} \, dx = \frac{1-x}{2-x}.
\]
Example: Gamma and Uniform

Suppose that $X$ has $\text{Gamma}(2, \lambda)$ distribution, and that given $X=x$, $Y$ has $\text{Uniform}(0,x)$ distribution.

**Question:** Find the joint density of $X$ and $Y$

**Solution:**

By definition of $\text{Gamma}(2, \lambda)$ distr., $f_X(x) = \begin{cases} \lambda^2 x e^{-\lambda x}, & \text{if } 0 < x; \\ 0, & \text{otherwise}. \end{cases}$

By definition of $\text{Unif}(0,x)$, $f_Y(y|X = x) = \begin{cases} \frac{1}{x}, & \text{if } 0 < y < x; \\ 0, & \text{otherwise}. \end{cases}$

By multiplication rule, $f(x,y) = f_Y(y|X = x)f_X(x) = \begin{cases} \lambda^2 e^{-\lambda x}, & \text{if } 0 < y < x; \\ 0, & \text{otherwise}. \end{cases}$

**Question:** Find the marginal density of $Y$.

**Solution:** $f_Y(y) = \int_0^\infty f(x,y) \, dx = \int_y^\infty \lambda^2 e^{-\lambda x} \, dx = \lambda e^{-\lambda y}$.

So $Y \sim \text{Exp}(\lambda)$. 
Condition Distribution & Expectations

Discrete | Continuous

- Conditional distribution of $Y$ given $X=x$:

\[ P(Y \in B | X = x) = \sum_{y \in B} P(Y = y | X = x) \quad \text{for discrete } X \]

\[ P(Y \in B | X = x) = \int_B f_Y(y | X = x) \, dy \quad \text{for continuous } X \]

- Expectation of a function $g(X)$:

\[ E(g(Y) | X = x) = \sum_{\text{all } y} g(y) P(Y = y | X = x) \quad \text{for discrete } X \]

\[ E(g(Y) | X = x) = \int g(y) f_Y(y | X = x) \, dy \quad \text{for continuous } X \]
Independence

Conditional distribution of $Y$ given $X=x$ does not depend on $x$:

$$P(Y \in B | X=x) = P(Y \in B).$$

Conditional distribution of $X$ given $Y=y$ does not depend on $y$:

$$P(X \in A | Y=y) = P(X \in A).$$

The multiplication rule reduces to:

$$f(x,y) = f_X(x)f_Y(y).$$

Expectation of the product is the product of Expectations:

$$E(XY) = \int \int xy \, f_X(x) \, f_Y(y) \, dx \, dy = E(X) \, E(Y).$$
Example: \((X,Y)\) Uniform on a Unit Circle.

\[
f(x, y) = \begin{cases} \frac{1}{\pi}, & \text{if } x^2 + y^2 < 1; \\ 0, & \text{otherwise.} \end{cases}
\]

Question: Are \(X\) and \(Y\) independent?

Solution:

\[
f_X(x|Y=y) = \begin{cases} \frac{1}{2\sqrt{1-y^2}}, & \text{if } -\sqrt{1-y^2} < x < \sqrt{1-y^2}; \\ 0, & \text{otherwise.} \end{cases}
\]

So the conditional distribution of \(X\) given \(Y=y\) depends on \(y\) and hence \(X\) and \(Y\) cannot be independent. However, note that for all \(x\) and \(y\)

\[
E(X|Y = y) = E(X) = 0 \text{ and } E(Y|X=x) = E(Y) = 0
\]
Average Conditional Probabilities & Expectations

Discrete

• Average conditional probability:

\[
P(B) = \sum_{\text{all } x} P(B|X = x)P(X = x)
\]

\[
P(Y = y) = \sum_{\text{all } x} P(Y = y|X = x)P(X = x)
\]

Continuous

• Average conditional probability:

\[
P(B) = \int P(B|X = x)f_X(x) \, dx
\]

\[
f_Y(y) = \int f_Y(y|X = x)f_X(x) \, dx
\]

• Average conditional expectation:

\[
E(Y) = \sum_{\text{all } x} E(Y|X = x)P(X = x)
\]

\[
E(Y) = \int E(Y|X = x)f_X(x) \, dx
\]