Introduction to Probability
Conditional distributions of random variables

Conditional probability, law of total probability

Recall the conditional probability of an event E, given F:

$$P(E \mid F) = \frac{P(EF)}{P(F)}$$

and the Law of Total Probability: if $\{E_1, ..., E_n\}$ partitions Ω,

$$P(F) = \sum_{i=1}^{n} P(FE_i) = \sum_{i=1}^{n} P(F \mid E_i)P(E_i).$$

We will now translate these to the language of random variables.

Conditional probability, conditional expectation

For discrete RVs, the **conditional PMF** of $X = x$, given the event F, gives the probability of the event $E(x) = \{X = x\}$ under the evidence F.

$$p_{X|F}(x) = P(E(x) \,|\, F) = P(X = x \,|\, F) = \frac{P(\{X = x\} \cap F)}{P(F)}.$$

Conditional probability, conditional expectation

The **conditional expectation** of X, given the event F, is

$$E(X \mid F) = \sum_{x \in X(\Omega)} x \, p_{X|F}(x).$$

Note that we only need to sum over the values in $X(F)$, not $X(\Omega)$, since our conditional probability only gives values from F positive probability. However, we will write $X(\Omega)$ here for consistency.

Expectation, PMF via conditional expectations, PMFs

If $\{F_1, ..., F_n\}$ partitions Ω, then we can compute the "regular" unconditioned PMF as a sum of conditional PMFs, mirroring the Law of Total Probability:

$$p_X(x) = \sum_{i=1}^{n} P(X = x \mid F_i) P(F_i) = \sum_{i=1}^{n} p_{X|F_i}(x)\, P(F_i).$$

Expectation, PMF via conditional expectations, PMFs

Likewise, we can compute the "regular" unconditioned expectation as a sum of conditional expectations:

$$E(X) = \sum_{i=1}^{n} E(X \mid F_i)P(F_i) = \sum_{i=1}^{n} \left(\sum_{x \in X(\Omega)} x \, p_{X|F_i}(x) P(F_i) \right)$$

$$= \sum_{x \in X(\Omega)} x \left(\sum_{i=1}^{n} p_{X|F_i}(x) P(F_i) \right) = \sum_{x \in X(\Omega)} x \, p_X(x).$$

Conditional probabilities with random variables

For discrete RVs, the **conditional PMF** of $X = x$, given $Y = y$, uses the joint PMF and the marginal PMF of Y in the same way.

Consider the events $E(x) = \{X = x\}$ and $F(y) = \{Y = y\}$. Then

$$p_{X|Y}(x|y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}.$$

Conditional probabilities with random variables

Likewise, the conditional PMF of $Y = y$, given $X = x$, is

$$p_{Y|X}(y|x) = P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{p_{X,Y}(x, y)}{p_X(x)}.$$

Joint PMF via conditional PMF

We can recover the joint PMF of X and Y if we start with the conditional PMF:

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \implies p_{X,Y}(x,y) = p_{X|Y}(x|y)\, p_Y(y).$$

This joint PMF can be written even when $p_Y(y) = 0$, since for these y, $p_{X,Y}(x,y) = 0$ as well.

Bayes' Law with random variables

This means Bayes' Law works similarly: in the general setting,

$$P(E \mid F) = \frac{P(F \mid E)P(E)}{P(F)}.$$

In the specific case of random variables,

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

Also, notice how $p_Y(y)$ uses the law of total probability:

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{\sum_{z \in X(\Omega)} p_{Y|X}(y|z)p_X(z)}.$$

Expectation, PMF via conditional expectations, PMFs

If $\{F(y)\}_{y \in Y(\Omega)}$ partitions Ω, then we can compute the "regular" unconditioned PMF as a sum of conditional PMFs, mirroring the Law of Total Probability:

$$p_X(x) = \sum_{i=1}^{n} P(X = x \mid Y = y) P(Y = y) = \sum_{y \in Y(\Omega)} p_{X|Y}(x|y)\, p_Y(y).$$

Expectation, PMF via conditional expectations, PMFs

Likewise, we can compute the “regular” unconditioned expectation as a sum of conditional expectations:

$$
\begin{aligned}
E(X) &= \sum_{y \in Y(\Omega)} E(X \mid Y = y) P(Y = y) \\
&= \sum_{x \in X(\Omega)} x \sum_{y \in Y(\Omega)} p_{X|Y}(x|y) p_Y(y) = \sum_{x \in X(\Omega)} x \, p_X(x).
\end{aligned}
$$

Expectation of a function of an RV

If $g : \mathbb{R} \to \mathbb{R}$ is a function, we can likewise compute the expectation of $g(X)$ using conditional expectations:

$$E(g(X) \mid F) = \sum_{x \in X(\Omega)} g(x)\, p_{X|F}(x),$$

$$E(g(X) \mid Y = y) = \sum_{x \in X(\Omega)} g(x)\, p_{X|Y}(x|y),$$

$$E(g(X)) = \sum_{y \in Y(\Omega)} E(g(X) \mid Y = y) P(Y = y).$$

Conditional probabilities with random variables: example

The experiment: flip a fair coin 4 times. Let

- X = number of H flipped on all four flips and
- Y = number of T flipped in the first three flips.

(a) What is the probability that $X = 3$?

(b) What is the probability that $X = 3$, given that $Y = 1$?

(c) Say we play a game:
you win \$1 for each H and lose \$1 for each T.
What is the probability you "win" the game, i.e. you finish with positive winnings?

(d) What are your expected winnings on the game in (c)?

Conditional probabilities with random variables: example

The experiment: flip a fair coin 4 times. Let

- X = number of H flipped on all four flips and
- Y = number of T flipped in the first three flips.

(a) $P(X = 3) = ?$

(b) $P(X = 3 \mid Y = 1) = ?$

(c) X = number of H flipped, so $4 - X$ = number of tails. Hence, your winnings on this game are

$$W = X - (4 - X) = 2X - 4.$$

Thus, the probability you "win" the game is $P(W > 0) = ?$

(d) $E(W) = ?$

Conditional probabilities with random variables: example

The experiment: flip a fair coin 4 times. Let

- X = number of H flipped on all four flips and
- Y = number of T flipped in the first three flips.

(a) $P(X = 3) = ?$

Note that $X \sim Bin(4, 1/2)$.

Hence,

$$P(X = 3) = \binom{4}{3} \left(\frac{1}{2}\right)^4 = \frac{1}{4}.$$

(This is from the marginal PMF for X.)

Conditional probabilities with random variables: example

The experiment: flip a fair coin 4 times. Let

- X = number of H flipped on all four flips and
- Y = number of T flipped in the first three flips.

(b) $P(X = 3 \mid Y = 1) = ?$

To find $P(X = 3 \mid Y = 1)$, we need the joint PMF for (X, Y).

We'll calculate it by writing out all $2^4 = 16$ possible cases.

We can first compute the marginals: we know

$X \sim Bin(4, 1/2)$, and it should also be clear that $Y \sim Bin(3, 1/2)$.

Conditional probabilities with random variables: example

	Y=0	Y=1	Y=2	Y=3	$p_X(x)$
X=0					$\frac{1}{16}$
X=1					$\frac{4}{16}$
X=2					$\frac{6}{16}$
X=3					$\frac{4}{16}$
X=4					$\frac{1}{16}$
$p_Y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

Conditional probabilities with random variables: example

	Y=0	Y=1	Y=2	Y=3	$p_X(x)$
X=0	0	0	0	TTTT	$\frac{1}{16}$
X=1	0	0	TTHT, THTT, HTTT	TTTH	$\frac{4}{16}$
X=2	0	THHT, HTHT, HHTT	TTHH, THTH, HTTH	0	$\frac{6}{16}$
X=3	HHHT	THHH, HTHH, HHTH	0	0	$\frac{4}{16}$
X=4	HHHH	0	0	0	$\frac{1}{16}$
$p_Y(y)$	$\frac{2}{16}=\frac{1}{8}$	$\frac{6}{16}=\frac{3}{8}$	$\frac{6}{16}=\frac{3}{8}$	$\frac{2}{16}=\frac{1}{8}$	1

Conditional probabilities with random variables: example

	Y=0	Y=1	Y=2	Y=3	$p_X(x)$
X=0	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$
X=1	0	0	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{4}{16}$
X=2	0	$\frac{3}{16}$	$\frac{3}{16}$	0	$\frac{6}{16}$
X=3	$\frac{1}{16}$	$\frac{3}{16}$	0	0	$\frac{4}{16}$
X=4	$\frac{1}{16}$	0	0	0	$\frac{1}{16}$
$p_Y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

Conditional probabilities with random variables: example

(a) Note that $X \sim Bin(4, 1/2)$. Hence, $P(X = 3) = \binom{4}{3}(\frac{1}{2})^4 = \frac{1}{4}$.

(b) $P(X = 3 \,|\, Y = 1) = \frac{p_{X,Y}(3,1)}{p_Y(1)} = \frac{\frac{3}{16}}{\frac{3}{8}} = \frac{1}{2}$.

(c) $P(2X - 4 > 0) = P(X > 2) = P(X = 3) + P(X = 4) = \frac{5}{16}$.

(d) $E(W) = E(2X - 4) = \sum_{j=0}^{4}(2j - 4)p_X(j)$

$$= -4\left(\frac{1}{16}\right) - 2\left(\frac{4}{16}\right) + 0\left(\frac{6}{16}\right) + 2\left(\frac{4}{16}\right) + 4\left(\frac{1}{16}\right) = 0.$$

Continuous conditional density, Bayes' law, joint density

The **conditional PDF** of X given Y is defined in a similar fashion as the discrete case.

If $f_{X,Y}(x,y)$ is the joint PDF of X and Y, and the marginal PDFs are $f_X(x)$ and $f_Y(y)$, then the conditional densities are

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)},\ f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_x(x)},$$

where the marginals are positive.

Bayes' law follows similarly:

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_x(x)}.$$

Continuous marginal density, law of total probability

We can recover the joint density from a conditional density as in the discrete case:

$$f_{X,Y}(x,y) = f_{Y|X}(y \mid x) f_X(x).$$

The marginal PDF of X can be computed as before, by integrating over all possible conditions for Y:

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

Thus, the law of total probability here gives us

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x \mid y) f_Y(y)}{\int_{-\infty}^{\infty} f_{X|Y}(x \mid z) f_Y(z) dz}.$$

Conditional expectation of X given Y

The **conditional expectation** of X, given Y, is a random variable that can be considered a function of the random variable Y.

For fixed $y \in Y(\Omega)$,

$$E(X \mid Y = y) = \sum_{x \in X(\Omega)} x \, p_{X|Y}(x|y).$$

Thus, we define the function $h(y) = E(X \mid Y)(y)$ by

$$E(X \mid Y)(y) = E(X \mid Y = y).$$

Conditional expectation of X given Y

We can use the conditional expectation function to compute results about X if you have the scenario where you can only access X through observations of Y.

For example, we can compute $E(X)$ using only conditioning on Y:

$$
\begin{aligned}
E(X) &= \sum_{x \in X(\Omega)} x \, p_X(x) \\
&= \sum_{x \in X(\Omega)} x \sum_{Y \in Y(\Omega)} p_{X|Y}(x|y) p_Y(y) \\
&= \sum_{Y \in Y(\Omega)} \left(\sum_{x \in X(\Omega)} x \, p_{X|Y}(x|y) \right) p_Y(y) \\
&= \sum_{Y \in Y(\Omega)} E(X \mid Y)(y) p_Y(y) = E(E(X \mid Y)).
\end{aligned}
$$

Conditional variance of X given Y

Likewise, we can compute the variance of X using only conditioning on Y: defining the **conditional variance** of X given $Y = y$ by

$$Var(X \mid Y = y) = E[(X - E(X \mid Y = y))^2 \mid Y = y],$$

we have the conditional variance as a function of y:

$$v(y) = Var(X \mid Y)(y) = Var(X \mid Y = y),$$

and we can compute the following identity:

Conditional Variance Formula:

$$Var(X) = E[Var(X \mid Y)] + Var(E(X \mid Y)).$$

Independence of X and Y via conditional, joint PDF

Two continuous random variables X and Y are called **independent** if conditioning one with the other yields no change in PDF. That is, $X \perp Y$ if

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

for every $x, y \in \mathbb{R}$. Equivalently, $X \perp Y$ if, for every $x, y \in \mathbb{R}$,

$$f_{Y|X}(y \,|\, x) = f_Y(y).$$

We can state this in terms of the joint PDF, as we have before: $X \perp Y$ iff, $\forall x, y \in \mathbb{R}$, the joint PDF factors into marginals:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

Wald's Identity

Theorem

Wald's Identity:

Let $X_1, X_2, X_3, \ldots$ *are IID RVs with* $E(X_1) = \mu < \infty$.
Let $N \in \{0, 1, 2, ...\}$ *be a RV with* $E(N) = \nu < \infty$ *and* $N \perp X_i \ \forall i$.
Set

$$S_N = \sum_{i=1}^{N} X_i.$$

Then,

$$E(S_N) = E(N) \cdot E(X_1) = \nu\mu.$$

This theorem allows us to easily calculate the expectation of the sum of a random number of random variables, as long as all the variables (and the counting RV N) are all independent.